DPYSL2 and mTOR Signaling in Schizophrenia

By

Xuan T. Pham

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy.

Baltimore, Maryland
March, 2016

Abstract

Background

Located in a schizophrenia susceptibility locus on chr8p21, *DPYSL2* has been implicated in schizophrenia by numerous linkage and association studies; however research on its biology in relation to the risk for SZ has been sparse. *DPYSL2* encodes CRMP2, a protein that functions in axon growth and maintenance. A first targeted sequencing study of this gene revealed several non-coding variants that altered expression in transient transfection assays. Most interestingly, a polymorphic CT di-nucleotide repeat (DNR) variant located in the 5'-untranslated region of *DPYSL2* is associated with SZ risk and disrupts regulation of the gene by mTOR signaling pathway. Given the biological relevance of *DPYSL2* to schizophrenia, the aim of my thesis, composed of three separate investigations, was to further explore the molecular mechanisms on which variations in *DPYSL2* contribute risks for SZ.

Methods

I extended the sequencing screen of this gene on a larger scale using a high-throughput pooling strategy, with the goal to identify rare functional variants in *DPYSL2*. The targeted regions for sequencing included the 14 exons and 27 conserved noncoding regions in and around this gene. To expand our knowledge on the polymorphic DNR

variant, we performed protein-RNA interaction studies that identified differential binding of mTOR effectors to the DNR alleles. And to study the DNR variant in its native genomic context, we used CRISPR/cas9 genome editing to create two isogenic cell lines that differ only at the DNR locus. To study loss of function of *Dpysl2 in vivo*, we created a conditional knockout mouse model of *Dpysl2* to test the hypothesis that loss of function of *Dpysl2* alters axonogenesis, perturbing morphological and behavioral phenotypes.

Results

From the extended sequencing screen of this DPYSL2, we identified a total of 286 variants, of which 66 were novel, and 8 were synonymous. The resulting comprehensive variant catalogue of this gene showed a modest enrichment of rare variants in SZ patients with no functional biases. Because the enrichment was small (1.28 fold-change) and not variant-specific, we decided to pursue the aforementioned DNR variant instead. In protein-RNA interaction studies, we observed differential bindings of proteins to the DNR alleles. Specifically, the longer risk allele lost binding to ELAVL4/HuD, a neuronal ribosomal binding protein regulated by the mTOR pathway. Additionally, the risk allele also had decreased binding to two mTOR effectors, 4E-BP and eIF4E. In the CRISPR-edited cells that carry the homozygous DNR alleles, we showed that the longer CT repeat leads to significant reduction of CRMP2 levels,

accompanied by striking shortening of cellular projections, and significant transcriptome changes that link this transcript back to SZ. We also successfully created a conditional *Dpysl2* knockout mouse, which shows interesting phenotypic changes in preliminary behavioral tests. The results will be confirmed and extended with additional testing, and this mouse model has the potential to be a powerful tool in the study of SZ.

Conclusion

Together my work contributes a significant body of evidence that link *DPYSL2* and mTOR to risk for schizophrenia. My results support the functional importance of the DNR, suggest a role for mTOR signaling, and further implicate *DPYSL2* as a schizophrenia susceptibility gene.

Advisor: Dimitrios Avramopoulos

Reader: Mikhail Pletnikov

Acknowledgements

The saying goes, “It takes a village to raise a child,” and so it also takes a village to raise a PhD Human Genetics student. My thesis dissertation is a product of great passion and labor that was made possible only with the guidance and support of many people. Here, I would like to sincerely acknowledge those people to whom I am eternally grateful.

Professional

To say that my entire graduate career would not be possible without the help of my thesis advisor and mentor, Dr. Dimitri Avramopoulos, would be a severe understatement. I’ve known Dimitri since 2008, when I had the great fortune of a summer undergraduate internship in his lab. The exposure to his lab and his mentorship made a great impression on me, and undoubtedly influenced my decision to do research in his lab again for the past five years. In all the times I’ve known Dimitri, he has always been calm, insightful, and resourceful. Even through the tightest experimental messes, Dimitri has never lost his cool and has never failed to offer practical alternatives and solutions to my problems. His insight and quick thinking has saved my thesis project and sanity countless times. Dimitri is the eternal optimist, as he is always able to see the glass-half-full in any situation. I don’t think

either of us ever thought our careers would be so inextricably linked after that short summer, but I am grateful for having had such a thoughtful mentor.

I would also like to thank the members of the Avramopoulos lab, both past and present. Our lab manager, Ruihua Wang, has provided me countless technical guidance ever since my summer rotation in 2008. She has helped troubleshoot many of my PCR and western blots, which always seemed to work better in her expert hands. Of note, I'd like to thank my wonderful lab-mate, colleague, and friend, Nicole Eckart. Having her in the lab to commiserate and celebrate throughout this journey has been wonderful. Nicole is one of the kindest people I've ever met, and I will truly miss our random lengthy chats in lab.

I am lucky that my lab family extends to the Valle lab, a constant source of knowledge, advice, and antics. In particular, Dr. David Valle continues to be one of the most astute scientist and mentor for not only our lab, but also the entire Human Genetics Program. He leads by example and pushes us all to accept only the best from ourselves. The Valle lab rotates many students and fellows, many have helped me out on numerous occasions; however, through any changes in the lab, I can always count on the help and support of Gary Steel and Cassandra Obie. Gary provided me with invaluable help with the mouse project, and

Cassandra remains the resident expert to whom I turn to for cell culture advice.

The Avramopoulos/Valle labs function with the vital help from Sandy Muscelli. Sandy works tirelessly, taking care of our lab, the program, and all the students, making sure all aspects of the IGM run smoothly. As busy as she is, Sandy has always made time to help me find answers to my problems, as she does with every single student in the program. Her love for the Human Genetics program and its members extends beyond the deepest depths of her glorious candy bowl.

I would like to thank the Human Genetics program as a whole, for the fantastic opportunity to study genetics at Johns Hopkins. Special thanks to Dr. David Valle, Dr. Kirby Smith, Dr. Andy McCallion, and Sandy Muscelli, for keeping the program running smoothly for the current students, and improving the program for the next generation of students.

I'd like to also thank my thesis committee members, Dr. David Valle, Dr. Andy McCallion, Dr. Kathy Burns, and Dr. Mikhail Pletnikov for their sound guidance throughout my project. The progress of my dissertation was largely shaped by their direction and insight.

Personal

To study human genetics is to be constantly reminded of the importance of family. Throughout education, I have always been conscious that my family made my journey possible. We emigrated from Vietnam when I was 9, as my parents were in search of better futures for their family and children. In particular, my father has always been the champion for education. Even when we had little to nothing, my education was always a priority, as my father believed knowledge would allow us to better face life's many challenges. To him, I owe my entire education and lifelong passion for learning. My mom epitomizes kindness and compassion, the strengths that carried our family through the toughest of times. I can only hope to emulate her quiet strengths to all the people I encounter in life. My sisters have traveled the roads ahead of me and been my protectors. They've been great sources of inspiration as successful businesswomen and devoted mothers of their families. I am grateful for their support as I ventured into a field that our family has never known before.

Last but never least, I would like to express my deepest thanks to my husband, Huailing, who is my biggest supporter and confidant. He has literally moved borders, jobs, and homes to be by my side as I pursued an education, and countless many other hobbies. Throughout this journey, I could always count on his immense kindness, patience,

and optimism as I navigated through some of the biggest changes in my life. In addition, he provided me with unrivaled technical support and always made sure my hard-earned data were diligently backed up in at least three separate locations. He pushes me to be my best and encourages me to move forward even when I didn't know I could. This experience is so much sweeter with him by my side.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Schizophrenia (SZ)

Schizophrenia (MIM #181500), henceforth referred to as "SZ," is a common and disabling neuropsychiatric disorder that can alter a person's thought and cognitive awareness. The disorder affects as much as ~1% of people worldwide, and causes a wide range of cognitive symptoms in early adulthood that can severely impair many facets of ordinary life [1]. In addition, SZ causes a huge economic burden, costing in excess of $62 billion in 2002 for the United States alone [2]. Despite the clinical, social, and financial impacts of SZ, the etiology of the disorder remains elusive though evidence suggest a strong genetic component [3]. Twin and adoption studies suggest that SZ is highly heritable, with concordance rates ranging from 40-65% for monozygotic twins and 5-15% for dizygotic twins [4]. Furthermore, adoption studies implicate shared environment as contributing risk factors for developing SZ [5,6].

DPYSL2 as a SZ susceptibility gene

Genome wide linkage scans and candidate gene association studies have revealed a number of SZ susceptibility genes [7,8], including *DTNBP1* (MIM *607145), *DISC1* (MIM *605210), *NRG1* (MIM *142445), and *NRG3* (MIM *605533). In particular, through one of the largest SZ SNP-based

linkage studies to date, our group identified region 8p21-22 as a candidate SZ susceptibility locus (p=0.0001) [9]. Our result was replicated by many other groups in different population [10–12], and we later performed a candidate gene-based association study targeting the 8p21 region in the Ashkenazi Jewish population [13]. We found the strongest linkage evidence over SNP rs12155555, located ~5 kb upstream of *DPYSL2*, and the strongest association signals from the follow-up fine mapping were located within or flanking the *DPYSL2* gene [13].

The dihydropyrimidinase-like 2 gene (*DPYSL2*) encodes a 62kDa cystolic protein best known as collapsin response mediator protein 2 (CRMP2), a member of a 5-gene family named for their involvement in axonal growth cone collapse. Mammals express three different transcripts of *DPYSL2*: A, B and C, each differing in the alternative first exon. Nearly all published literature has centered on the B transcript, which is also the transcript of this investigation, and is henceforth referred to simply as *DPYSL2*. Located at 8p21.2 (26.5 Mb), *DPYSL2* is primarily expressed in the central nervous system, though it is also found at lower levels in the peripheral nervous system [14]. This gene is reported to have vital functions in neuronal development [15–18], with dysfunctions resulting in neurodevelopmental abnormalities, which may be a contributing factor in the pathogenesis of SZ. Knockout mutants of CRMP2 display abnormal dendritic formation in hippocampal cells, suggesting that CRMP2 is essential for proper dendritic organization [19].

Further, aberrant expression of *DPYSL2* has been reported in multiple areas of the SZ postmortem brain, including the hippocampus [20], the frontal cortex [21], the anterior cingulate cortex [22], and the dorsolateral prefrontal cortex [23]. The CRMP2 protein interacts with other proteins to stabilize microtubules, promote neurite outgrowth, and modulate signaling processes in the central nervous system [24–27]. In cell culture studies, altering the expression of Crmp2 significantly affected the axonal growth of rat hippocampal neurons [19,28]. Most recently, a study showed downregulation of Crmp2 protein in the prefrontal cortex and hippocampus of rats that were stressed prenatally [29]. The study also identified 2 SNPs that were associated with schizophrenia susceptibility, further suggesting a role for *DPYSL2* in this disease [29].

Initial sequencing study of *DPYSL2*

In a previous study [30], we sequenced 137 SZ patients and similar number of controls and identified 120 variants, some of which were significantly associated with SZ. Several of these are in conserved non-coding regions (cNCRs) and show sex-specific risks for SZ with odds ratio of 2 – 2.5 in one sex but not the other. Additionally, a few variants in the promoter region and intron 1 of *DPYSL2* showed evidence of directing expression patterns in the central nervous system when expressed in the zebrafish, pointing to the regulatory potential of these variants. Most interestingly, the study identified a polymorphic CT di-nucleotide repeat

(DNR) located in the 5'-untranslated region (UTR) of this gene [30]. The DNR disrupts a reported 5'-terminal oligopyrimidine tract (5'-TOP), a consensus sequence marked by a Cytosine that preceeds a continuous tract of pyrimidines [30–32]. Translation of 5'-TOP mRNAs is thought to be regulated by the mTOR pathway [33,34]. In their investigation, Liu et al. reported 11 CT repeats (11DNR) in the common allele and 13 CT repeats (13DNR) for the alternative allele, with the risk allele reducing translation by 3-fold in transient expression systems in cells and in polysome profiling experiments. Further, we reported a dose-dependent response of the two DNR alleles to Rapamycin, an allosteric inhibitor mTOR, suggesting regulation of *DPYSL2* by the mTOR pathway. This result delineates the DNR as a functional genetic variant and suggests a mechanism for coupling SZ risk to environmental variables through mTOR signaling.

Thesis summary

Given the biological relevance of *DPYSL2* to schizophrenia, the aim of my thesis was to further explore the molecular mechanisms on which variations in *DPYSL2* contribute risks for SZ. To this end, my thesis is composed of three separate but related investigations. First, I tested the hypothesis that SZ patients carry more rare functional variants than the controls variation in and around *DPYSL2*. Specifically, I extended the sequencing screen of the exons and conserved sequences around this

gene on a larger scale, in a more homogeneous population, and with a more high-throughput sequencing strategy. In the resulting comprehensive variant catalogue of this gene, we found a modest enrichment of rare variants in SZ patients with no functional biases. Because the enrichment was small (1.28 fold-change) and not variant-specific, we decided to not follow-up any individual variants from this screen.

Second, I expanded our knowledge on the polymorphic CT DNR variant, showing that it perturbs mTOR signaling via differential binding of different mTOR effectors. We created isogenic cell lines that differ only at the DNR locus, and showed the longer CT repeat leads to significant reduction of CRMP2 levels, accompanied by striking shortening of cellular projections, and significant transcriptome changes that link this transcript back to SZ. The results from this study strengthen the evidence that this variant is likely a risk factor for SZ, with the risk allele altering cellular phenotypes at the protein, morphology, and transcriptome level.

Finally, we created a knockout mouse model of *Dpysl2* to test the hypothesis that loss of function of *Dpysl2* alters axonogenesis, perturbing morphological and behavioral phenotypes. To my knowledge at this time, no such mouse model exists for this gene in the C57BL/6 background. Our preliminary behavioral tests showed possible

phenotypic changes in the knockouts, the results of which could be confirmed and extended with additional testing. The conditional *Dpysl2* knockout mouse has the potential to be a powerful tool in the study of SZ and genetics in general.

Together my work contributes a significant body of evidence that link *DPYSL2* and mTOR to risk for schizophrenia. My results support the functional importance of the DNR, suggest a role for mTOR signaling, and further implicate *DPYSL2* as a schizophrenia susceptibility gene.

Figures: Chapter 1

Figure 1.1: Schematic of the 5'-TOP tract of DPYSL2

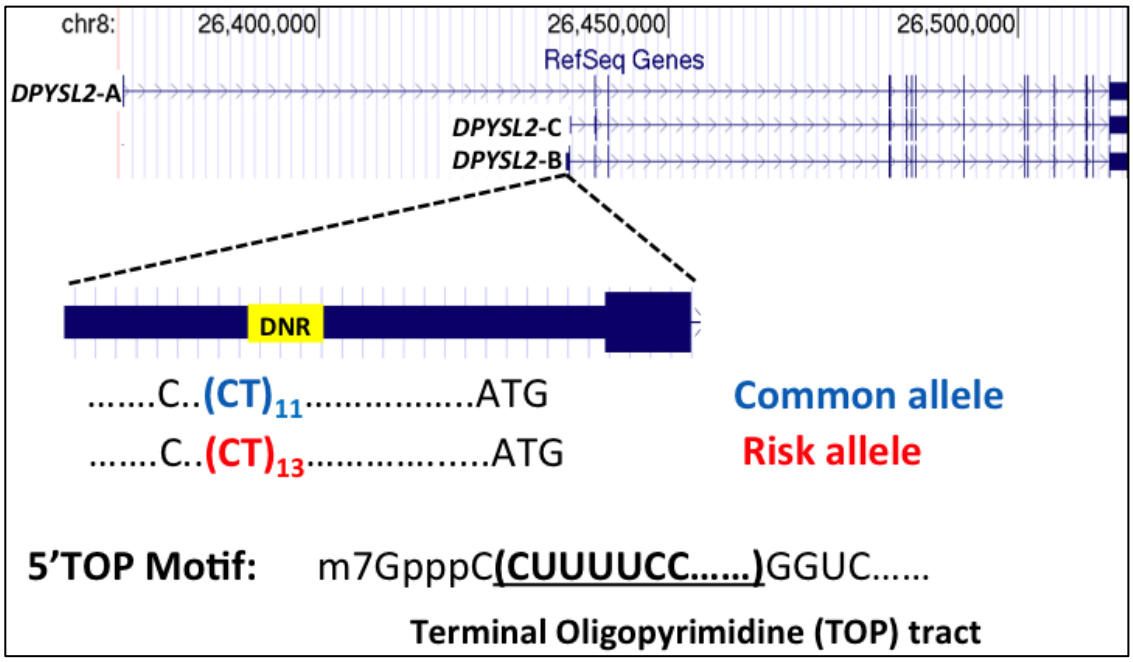

Figure 2.1: The polymorphic CT di-nucleotide repeat (DNR) variant is located in the 5' UTR of *DPYSL2*, which is also part of a reported 5'-terminal oligopyrimidine tract (5'-TOP), a consensus sequence defined as having a Cytosine preceding an uninterrupted stretch of pyrymidines.

Chapter 2: Next Generation Sequencing Screen of *DPYSL2*

Abstract

Linkage and association studies have repeatedly implicated *DPYSL2* on chr8p21 as a candidate gene for SZ; however, no causal variants have been identified. *DPYSL2* encode CRMP2, which is important in axonal growth, and dysfunction of this gene may result in neurodevelopmental abnormalities. In a first targeted sequencing study of this gene, we previously identified several non-coding variants in and around *DPYSL2* that altered expression in transient transfection assays. In order to capture more rare and functional variants in *DPYSL2*, we extended the sequencing screen by using a high-throughput strategy in a homogenous AJ population. We included all known *DPYSL2* exons and alternative exons as well as conserved non-coding regions within and around the gene. Our screening strategy identified 286 variants, of which 97% are non-coding and 24% are unannotated. With a comprehensive catalog of genetic variation at this locus, we found a modest enrichment of multiple rare variants in SZ cases compared to controls, though the variants did not show regulatory biases. In addition, two synonymous rare variants did not show exonic splicing potential in carrier lymphoblasts. Our results suggest variants at this locus could contribute

risk for SZ, though additional work is needed to follow up the specific variants that are functional and driving the risk.

Introduction

Our group was among the first to report strong linkage evidence for region 8p21 (27.61 Mb; 1-LOD interval 21.37-29.36 Mb) to SZ in both European Caucasian (EUC) and Ashkenazi Jewish (AJ) populations [9]. Since then, this linkage peak has been replicated by multiple groups in different populations, all converging on region 8p21 as a possible SZ susceptibility locus (SSL) [7,10–12,35,36]. Our group has investigated this 8p21 region in two ways: first, a family-based association study in 106 of our 8p-linked EUC families using 1402SNPs distributed across a 4.4 Mb interval (24.8-29.2 Mb) centered on the peak; and second, an independent case-control association study in a genetically more homogeneous AJ sample (709 cases, 1547 controls). In the EUC all SNPs with nominally significant associations contained in an 2.8Mb region (25.7-28.5 Mb) were selected for further study using 970 SNPs in our AJ case/control study. The strongest signal in the latter was for rs12155555 (OR = 0.66; CI: 0.53-0.81, p=0.00006) located about 3 Kb 3' of *DPYSL2* and the 9 SNPs with best p values are all located in the 256 Kb interval between the telomeric (*PNMA2*) and centromeric (*ADRA1A)* neighbors of *DPYSL2* [13].

To search for possible causative variants driving the linkage and association in this region, we have previously conducted targeted sequencing of the 14 exons and 27 conserved non-coding regions

(cNCRs) in and around *DPYSL2*. This was done in a small sample set (48 EUC and 89 AJ SZ probands, and 96 EUC and 56 AJ controls) [30]. This study identified 120 sequence variants, 38 of which were not in dbSNP (build 129); 30 of these had a MAF <1%. Of 10 coding variants, 9 were synonymous and one was a novel missense allele, A267S, present in a single EUC SZ proband. Several of these are in conserved non-coding sequences (cNCS) in the promoter region and intron 1 regions, which showed direct appropriate CNS expression pattern when expressed in zebrafish. Several of the human *DPYSL2* variants show sex-specific risks for SZ with OR of 2-2.5 in one sex but not the other. All of the risk alleles cause reduced expression in transient transfection assays in 293 cells and mouse primary cortical neurons. Most importantly, we identified a polymorphic dinucleotide variant (5'DNR) in the 5'UTR of *DPYSL2* that disrupts a 5'TOP sequence. The risk allele of this variant reduces translation by 3-fold in transient expression systems in cells and in polysome profiling experiments [30]. Taken together, these results add further evidence of significant association of SZ with SNPs in and around *DPYSL2*.

Spurred by the preliminary sequencing results, which were done in a small cohort of patients and controls, we embarked on a task to extend the sequencing efforts beyond what we previously reported. Specifically, our goal was to screen *DPYSL2* on a larger scale in a more homogeneous population using a more high-throughput sequencing strategy, with the

idea that this will allow us to capture more rare and functional variants in this gene. We hypothesized that SZ patients carry more of these rare, functional variants than the controls. Furthermore, by expanding our search for causative variants, we expected this would provide insight into how alterations in *DPYSL2* and its protein product, CRMP2, contribute risk for SZ.

To do this, we performed next-generation sequencing (NGS) of *DPYSL2* in 768 total of Ashkenazi Jewish patients and controls, targeting the same regions as the previous study, which included 14 exons and 27 conserved noncoding regions. Using a pooling strategy and next-generation sequencing, my screen identified 286 single nucleotide variants, the majority of which were non-coding. We tested this catalogue of variants for enrichment of rare functional variants between the two groups, and found a small, statistically significant enrichment for the SZ cases. Specifically there was about 1.3 fold more rare alleles in cases than in controls when the allele frequency less than 1%. We also explored whether variants that are enriched in SZ cases had more regulatory features, and found no significant overlap compared to those found in controls. In addition, we followed up two synonymous variants, which were predicted to alter binding of splicing factors. However, in testing *DPYSL2* splicing pattern of carrier vs. non-carrier lymphoblasts, we found the variants do not alter splicing of the exons in *DPYSL2*. Our results indicate that *DPYSL2* potentially harbors rare variants that occur

more often in SZ cases; however, the significance of this weren't striking enough to pursue follow-up of variants for individuals in the pooled screen.

Materials and Methods

Samples

The sample set in this study included 384 unrelated Ashkenazi Jewish (AJ) cases and 384 AJ controls, which were from the extensively phenotyped sample assembled and characterized by Dr. Pulver and her team [37]. All SZ cases where diagnosed based on DSM-IV criteria for schizophrenia or schizoaffective disorder [36]. All controls were screened for negative history of psychotic episodes, psychiatric hospitalization, and other SZ-relevant symptoms. Samples were collected with informed consent under the approval of the Johns Hopkins Institutional Review Board.

Identification of Regions of Interest

At the time of my sequencing studies, there was little evidence about *DPYSL2A* or *DPYSL2C*, so I focused exclusively on the B isoform, excluding the alternative first exons of the A and C isoforms. The regions of interest included the 14 known *DPYSL2* exons as well as 27 conserved non-coding regions (cNCRs) in and around the gene, as described

previously [30]. The cNCRs were selected on the basis of sequence conservation and computational prediction of enhancer elements [30].

Pooling and Next-Generation Sequencing

We used a pooling and next generation sequencing strategy [38] to quickly and cost-effectively sequence all the regions of interest in *DPYSL2.* DNA concentrations were obtained for each sample using a PicoGreen dsDNA assay. The samples were normalized and combined into pools, each containing 32 samples. This yielded an even 12 pools for cases and 12 pools for controls. Next, the 14 exons and 27 cNCRs were PCR amplified, which resulted in 46 high quality amplicons that covered 36.9 kb of *DPYSL2* (Table 2.1). The amplicons from the same pool were combined at equal molar ratio, resulting in 24 pools with 46 amplicons per pool. The pooled amplicons were ligated to avoid "end" enrichment effects, blunted and randomly fragmented using NEBNext dsDNA Fragmentase (NEB # M0348S). This fragmented DNA was then used to generate 24 indexed libraries using Illumina's TruSeq DNA Sample Preparation v2 Kit (Illumina # FC-121-2001, 2002). Indexed libraries were quantified and pooled together at equal molar ratio. The indexed libraries were run in one lane on our HiSeq 2000 genome analyzer in our genetics core facility to acquire 75 bp paired end reads. (Fig. 2.1)

Sequencing analysis

The paired reads were treated as single reads due to the nature of this experiment. The sequences were aligned and indexed to the reference genome (GRCh37/hg19) using Bowtie and SAMTools software. We used the Integrative Genome Viewer 2.0 (IGV) to visualize our reads and to identify variants. Because the frequency of one variable allele in a pool is very low (~1.56%), we improved sensitivity by also employing an automated system of variant selection in order to capture possible variants that might have been missed with IGV. We generated Perl scripts to sort through alignments and identify bases that were read as different nucleotides 0.5% of the time or more (expected alternative allele frequency in a pool with 1 heterozygote is 1.56%) taking into account information on strand bias. The total number of variants identified was 286 (Table 2.2).

Variant Filtering and Selection

From the total list of identified variants, we explored whether there was an enrichment of rare functional variants in the cases vs. the controls. To do this, we filtered the variants to select for those that had an MAF < 10% as reported by dbSNP. Variants that were novel and not catalogued by dbSNP were included in this set, as these were assumed to also be low in frequency by inference. Using the USCS Genome Brower

software, we also selected variants that were conserved and had regulatory features, defined by evidence of CpG islands, transcription factor binding sites, methylation features (H3K4me1, H3K4me3), acetylation features (H3K27Ac), and DNase hypersensitivity. These criteria filtered the 286 variants to a short list of 68 variants (Table 2.3). For each of the 68 variants, the estimated allele counts were calculated using the reported IGV allele frequencies for the case or controls, multiplied by 64 (the number of chromosomes per pool) and multiplied by 12 (the total number of pools for cases or controls). The total estimated allele counts for the cases and controls were compared for evidence of enrichment in either group using the chi-square test.

The 68 variants were then separated into those that were enriched in cases (n = 31) vs. those not enriched in cases (n = 37). For each subset, I compared the number of variants that overlapped with the regulatory features as described above.

Testing 2 synonymous variants for splicing potential

The synonymous variants (known: rs113199330 and novel at chr8: 26,484,146) were predicted by ESEFinder 3.0 to alter binding of serine/arginine-rich splicing factors (SRSF), which could affect the splicing patterns of DPYSL2 [39]. To test this potential, we assayed DPYSL2 transcript expressions in non-carrier vs. carrier lymphoblasts, which were identified by de-convoluting the pools containing the

variants. We obtained lymphoblast cultures from 6 Ashkenazi Jewish samples that were sequenced, 2 were carriers for the variants, and 4 were non-carriers, as well as 2 additional samples that were not AJ. From the lymphoblast pellets, we isolated RNA (Qiagen #74104) and synthesized cDNA using SuperScript III RT-PCR kit (Invitrogen #12574018). The cDNA was used as the template for PCR with primers specific at exon 3 and exon 6 of DPYSL2. The reaction was performed under standard PCR conditions, and products were visualized on 1% agarose gel as 480 bp bands.

Results

High coverage of mappable reads

Using the pooling strategy and next-generation sequencing, we received over 384 million 75 bp paired-end reads. The distribution of reads ranged from 9.7 to 27.3 million reads, with an average of 16.3 million reads across the 24 pools. In comparing the distribution of reads for cases vs. controls, the differences are not statistically significant ($p = 0.054$) (Fig. 2.2). The reads had > 80% alignment to the reference genome with an average allelic coverage of 381x.

Majority of variants identified are non-coding

Using both the manual and automated methods of variant classification, we identified a total of 286 single nucleotide variants (SNVs): 8 (~3%) are synonymous coding variants, and 278 (~97%) are noncoding (Table 2.2, Fig. 2.3). Of this, 65 variants (23%) were novel. The total number of SNVs identified in this study of 768 individuals (384 cases, 384 controls) is about ~2x more than the 120 variants identified in the first sequencing study of 192 samples [30]. In addition, these new results verified all the coding variants identified in the prior sequencing study.

Mild enrichment of rare and functional variants in SZ cases vs. controls

In comparing the variable allele counts between cases and controls, we found a small enrichment of rare variants in SZ cases vs. controls. Specifically, there was about 1.3 fold more rare alleles in cases than in controls when the allele frequency was less than 1% ($p = 4 \times 10^{-4}$). This was the highest fold change, as at an allele frequency of 10%, the enrichment was reduced to 1.08 fold ($p = 5 \times 10^{-4}$). The enrichment was non-significant when we compared between allele frequencies of 2% and 5% ($p > 0.1$) (Table 2.4). In addition, when we assessed whether variants

that are enriched in SZ cases have more regulatory features, we found no significant overlap compared to those found in controls.

Two synonymous variants do not alter DPYSL2 splicing

Our selection criteria included 2 synonymous variants, one known (rs113199330) and one novel (chr8: 26,484,146). These variants were analyzed for splicing potential using the Exonic Splcing Enhancer Finder (ESEFinder 3.0) [39], which reported alterations in loss and gain of several serine/arginine-rich splicing factors (SRSF) for both SNPs (Fig. 2.4A). In RT-PCR assays that tested splicing potential of the two synonymous variants in lymphoblast tissues, we identified a singular amplicons band at the expected size (~480 bp) for both carrier and non-carrier samples (Fig. 2.4B). This indicated that the two synonymous variants do not influence exon splicing of DPYSL2 in lymphoblasts.

Discussion

In this part of my thesis, our aim was to expand our search for causative variants on the rationale that this will provide insight into how alterations in *DPYSL2* contribute risk for SZ. We did this by performing a sequencing screen in a homogenous AJ population using a high-throughput strategy that quickly and cost-effectively captured all the regions of interest in the DPYSL2. Our pooling technique proved that

parallel sequencing is an effective method for deep coverage of a region of interest in many samples, as we obtained high coverage of every amplicon with thousand of reads covering the vast majority of the targeted regions. However, we noted that with 32 people per pool, the expected frequency of one variable allele per pool is very low (~1.56%) and close to the 1% due to background noise. To increase signal and better identify true positives, we recommend reducing the pool size to 16 people or less.

The result of this effort is a new comprehensive catalogue of variants in and around the DPYSL2 gene for well-researched homogenous AJ population. Overall, the paucity of synonymous variants (3% from total of identified variants) is remarkable and indicates that *DPYSL2* is highly conserved has strong biological relevance. By keeping the cases and controls separate, we could estimate whether certain variants are more common in one vs. the other group. This analysis yielded a small, statistically significant enrichment of rare variants in SZ cases vs. controls, a result that's interesting and supports our hypothesis that SZ patients are predisposed for more variants. However, the enrichment is small, is spread across all reads, and do not point to any one particular variant. Furthermore, when we looked to see if variants that are enriched in SZ cases have more regulatory features, we found no significant overlap compared to those found in controls. In analyzing two coding variants of this list that were predicted to alter

binding of splicing factors, we determined that these synonymous variants do not influence exon splicing of *DPYSL2* in lymphoblasts.

Overall, my results for this study are promising, as the catalogue of variants could have high utility with more advanced techniques such as massively parallel reporter assays, which can simultaneously test all of the variants for regulatory potential.

Tables: Chapter 2

Table 2.1: DPYSL2 sequencing primers

Primer ID	Sequence	Product (bp)
DPYSL2 E1 F	AATCGCTGCTCGTCTCTCTC	554
DPYSL2 E1 R	AGTTCGGTCCCTCTCCTTTC	
DPYSL2 E1 F'	ATTTGCATATCCCAGGATCG	364
DPYSL2 E1 R'	GACCGCAGATTTCCCAAGTA	
DPYSL2 E2 F	GGTTGCCACATACAGGCTTA	393
DPYSL2 E2 R	TGGTCACTCTCTGATGTTCCTG	
DPYSL2 E3 F	GCACAGAGCGGAGGATAGTT	376
DPYSL2 E3 R	GTGGCTCAGGACAGCAAGTT	
DPYSL2 E4 F	CGATTTTGAACCCAAGAAGC	425
DPYSL2 E4 R	ACAAATGGGATCTGGAAGCA	
DPYSL2 E8 F	CAGCCTCGCCTTCATCTTAG	421
DPYSL2 E8 R	ATACTAACAAGGCCGCAGCA	
DPYSL2 E9 F	TCTGCGAGATGAGCCTGATA	418
DPYSL2 E9 R	TTTGGTGAACAAGGGTCTCC	
DPYSL2 E10 F	GGGAGGGGATTCTGGATAAA	436
DPYSL2 E10 R	TGTCATTTCAAGGCCTCCAT	
DPYSL2 E11 F	CAACGCTCTTGACATCCATC	428
DPYSL2 E11 R	CCCAGCAAATTGCAAATCAT	
DPYSL2 E12 F	CAGGATCCCTGTCTCTGAGTCT	432
DPYSL2 E12 R	GCTCCCTCTTATCCCTGGAA	
DPYSL2 E13 F	CACACAACACCTGTCCACCT	435
DPYSL2 E13 R	CACGTTGGAGAGAAAGGGAAT	
DPYSL2 E14 F	GCACTTTTCCTCCTGAGCTG	551
DPYSL2 E14 R	GGTGGGTAGTCTTGGGTGTG	

DPYSL2 E5-7-1 F1	TGCTTCCTTTACCATTCTGTG	551
DPYSL2 E5-7-1 R1	AAGCTCTTGGGTGACACCTG	
DPYSL2 E5-7-2 F	CGGCACTCAGAACCTCCTTA	493
DPYSL2 E5-7-2 R	ACCAGGTCCCTCTGCTTTTC	
DPYSL2 E5-7-3 F	CAGGGCTTTCTTTTTCGTCA	533
DPYSL2 E5-7-3 R	ATTTTGCACAGCTCTGATGG	
DPYSL2 E5-7-4 F	GAACGATCGGCAGTGGTAAT	412
DPYSL2 E5-7-4 R	ATCTTCTGGCGAAACAGTCC	
3prime_cNCR1_1_F	CTGCACCCAGTGATGTGTGT	2218
3prime_cNCR1_2_R	TGCTGTTGTCCTCTGTGAAAA	
3prime_cNCR2_F	AGGCAGCACCAGGAAGAGT	517
3prime_cNCR2_R	CTGGCTTTTGATCTCGGTTC	
3prime_cNCR3_1_F1	GAGTTTGCCTCTGGGAAGTG	870
3prime_cNCR3_2_R	CTGTGGTCGACAGCACCTTA	
DPYSL2_3prime_new_cNCR8_F	CCTTCCTGCCAATTCTCTGA	644
DPYSL2_3prime_new_cNCR8_R	TCCCAAAGTGCTGGGATTAG	
DPYSL2_3prime_new_cNCR9_F	CCCCCACACCTTTCTTTCTT	590
DPYSL2_3prime_new_cNCR9_R	TACTAATACAAGCCTTATACCCTGTGC	
DPYSL2_3prime_new_cNCR10-1_F	GCACCCTATTCTCTCTCCTCCT	1613
DPYSL2_3prime_new_cNCR10-3_R	GGCCTTTGAAGGAAACCTCT	
DPYSL2 5'-1 F	CTCAGGGCTCTCATTTCCTG	715
DPYSL2 5'-6 R	GAGAGAGACGAGCAGCGATT	
DPYSL2 5'-2 F	ACGACCACCCCTCCATTGT	352
DPYSL2 5'-2 R	GGAAGGGGGTGGAAATAAAC	
DPYSL2 5'-3 F	CCAGCCACCTTTTGTAGCAC	503
DPYSL2 5'-3 R	AGCATCCCTCTGCTGTCATT	

DPYSL2 5'-4 F	CTGGGCAGTCAGGACACTCT	489
DPYSL2 5'-4 R	GAAATGAGAGCCCTGAGGAA	
DPYSL2 5'-5 F	AGATGGGGACTTGGATCCTC	608
DPYSL2 5'-5 R	GAGAGGTGGGAGCAAGACTG	
DPYSL2 U1-1 F	CTGCAGTGAGCCGAGATTTT	2410
DPYSL2 U1-6 R	TGCATTAACGGTCTTCCTACTG	
DPYSL2 U2-F1	CCCATCCATCAGCTCACAC	645
DPYSL2 U2-4 R	CCTAAGTCTCCCTCGCTGTG	
DPYSL2 U3-1 F	TACTGGAGCAGAAGCCCACT	1409
DPYSL2 U3-4 R	GACCACCAGGCTCAAGTGAT	
I3_cNCR1_F	AAATCACATGGTCCCAGCAC	530
I3_cNCR1_R2	CCCGTGTCCTTAGGAGAAATG	
I3_cNCR2_1_F	CCCCAAGCTGACCTTTCTCT	1234
I3_cNCR2_2_R	CACGCTCAGACAGAAAGCAC	
I3_cNCR3_F	GCATTCAAATGAGGCCAACT	3184
I3_cNCR3_R	TCAGGAGATGGAGAGGAGGA	
I3_cNCR4_1_F	GAGGGATACTTTGCATCTGGA	1938
I3_cNCR4_2_R	TCCCAACAGAATGTATCAGGAA	
DPYLS2_cNCS1_R	TGGGTGTCCTGACTGGCTAT	509
DPYLS2_cNCS1_F_OUT	TGACTTCTTGGGTGTCCTGA	
DPYLS2_cNCS2_F	GATGAACCGGAGTGGTTTGT	555
DPYLS2_cNCS2_R	TCCAAGCTGGAATTTGATCC	
DPYLS2_cNCS3_F	CTGTCCATCCCAGTTGGTTT	599
DPYLS2_cNCS3_R	CAGGCCTCTACCTCCACAAG	
DPYLS2_cNCS4_F_OUT	CTGTATCTCTGGGAGGTGCTG	350
DPYLS2_cNCS4_R_OUT	CCTTGTCTCCACCAGAGCTT	
DPYSL2_3prime_new_cNCR1_F	TGTGGTTGTTTCAGGGGTGT	625
DPYSL2_3prime_new_cNCR1_R	CCCAAGGCACACACTCTCTT	

DPYSL2_3prime_new_cNCR2_F	ATATTGCCTGGGCTGGTTTT	628
DPYSL2_3prime_new_cNCR2_R	CCCATCTTGGTGTGAGATGA	
DPYSL2_3prime_new_cNCR3_F	AGCCAAGATCGTGCCACTTA	457
DPYSL2_3prime_new_cNCR3_R	CCTGTGTGGTGCAGAGATGT	
DPYSL2_3prime_new_cNCR4_F	CCAAGAGGTGGCTTTTTGAT	455
DPYSL2_3prime_new_cNCR4_R	TACAGGCTGTCCCTGTCCTT	
DPYSL2_3prime_new_cNCR5_F	GAGGGTCTCCTTCCTGCTCT	521
DPYSL2_3prime_new_cNCR5_R	CTGGAATTGGGAGGGAAGTT	
DPYSL2_3prime_new_cNCR6-1_F	AGATCAGGCCTCCTCTACCC	1438
DPYSL2_3prime_new_cNCR6-3_R	CAGCGATATGCAAAATGAGG	
DPYSL2_3prime_new_cNCR7_F	CTCCCGAGTTGCTTGTGAAT	401
DPYSL2_3prime_new_cNCR7_R	TCAGGACTTTCAAGGCTGCT	
3prime_cNCR2_F	AAGCATTCTCTGTGGGAGTCA	399
3prime_cNCR2_R	TTCCCATCTTTTGCAGTAGGA	

Table 2.2: List of 286 SNVs identified through sequencing screen

Variant Coordinates	Ref / Alt. Allele	Known SNP
26366941	A/G	rs59632294
26367091	G/A	NO
26367158	G/A	NO
26371264	G/A	rs117982477
26371438	T/C	NO
26371543	C/A	rs17403251
26371729	C/G	NO
26371793	T/C	rs17055370
26372335	C/T	NO
26372351	A/G	rs149517015
26372354	G/A	NO
26372358	G/A	NO
26404953	T/C	rs192158859
26404960	G/A	rs9314324
26404986	A/C	rs73226141
26405024	G/A	rs13279683
26405185	G/T	rs78025765
26405344	A/T	NO
26405508	G/T	rs7831864
26405630	C/G	rs12548421
26405639	G/A	rs12543392
26405691	G/A	rs12543398
26405745	T/C	rs77020974
26405755	G/A	rs79293718
26405918	G/C	rs7823093
26406022	G/C	rs10110479
26406093	T/C	NO
26406095	A/G	rs182669480
26434583	A/G	NO
26434626	C/G	NO
26434795	C/G	rs431246
26434888	G/A	NO
26435116	C/G	rs367948
26435123	T/C	rs400181
26435129	C/A	rs76729423
26435213	C/G	NO
26435238	G/T	NO
26435271	T/C	rs445678
26435451	C/T	rs114900295

26435549	DNR	rs3837184
26435691	C/A	NO
26436000	T/A	NO
26436034	A/G	rs74917054
26436109	T/C	rs379266
26436221	G/A	NO
26436344	T/A	NO
26436349	A/G	NO
26436349	A/G	NO
26436351	G/A	NO
26436480	C/G	rs67474860
26436481	C/G	rs35321818
26436576	G/C	rs11781865
26436620	G/A	NO
26436770	G/A	rs2584184
26436783	G/A	rs62491888
26441264	T/C	NO
26441275	G/A	rs415524
26441477	G/A	rs11786691
26441529	G/A	rs188964405
26441562	G/A	rs408753
26447419	C/G	rs73226171
26447481	T/G	rs17055482
26447694	G/T	NO
26447739	G/A	rs140053560
26447767	C/T	rs143736793
26447790	A/G	rs62491913
26447791	T/C	rs113140418
26447830	T/A	rs68019818
26450473	A/G	rs7825468
26450631	T/C	rs141355118
26450858	G/A	rs146056239
26450941	C/T	rs77163697
26450950	G/A	rs140047324
26451098	G/C	rs17055487
26451134	C/G	NO
26452999	A/G	rs62491919
26453039	G/A	rs73226176
26453105	C/T	rs5029306
26453162	G/A	rs78609329
26453220	C/G	NO
26453468	G/T	rs11990271

26453486	T/A	rs114130812
26453655	G/A	rs17055489
26453830	C/T	rs12550663
26454048	C/T	rs17055491
26454316	C/T	rs17055492
26454793	C/T	rs76431903
26454911	G/A	rs17088245
26455093	C/T	rs185303505
26455100	C/T	rs62491920
26455145	C/T	rs181423952
26455242	G/A	rs113220440
26455403	G/A	rs77586745
26455418	T/C	NO
26455422	A/G	rs17406372
26455439	C/T	rs17055494
26455539	T/A	rs117245463
26455543	G/A	rs139911806
26455566	C/T	rs115825436
26455715	G/A	rs327232
26455800	T/C	rs7001738
26455881	C/T	NO
26455992	T/C	rs188591677
26455992	T/C	rs188591677
26466011	G/T	rs74787584
26466059	T/A	rs7820433
26466520	G/A	rs191369930
26466789	A/G	NO
26466803	C/T	rs116834469
26466860	C/A	rs11775967
26466877	G/A	rs62491954
26467541	C/T	rs146182569
26481666	C/T	rs113199330
26481771	C/T	rs327222
26481822	C/T	rs139123217
26484061	C/T	rs112899484
26484081	G/T	NO
26484146	C/T	NO
26484258	T/C	rs192142164
26484392	C/T	NO
26484404	G/A	NO
26484426	G/A	rs77752269
26484435	G/C	rs62493366

26484613	G/C	rs17322275
26484652	T/A	NO
26485062	C/T	rs327228
26485063	G/A	rs73678824
26485085	G/A	rs112094282
26485212	C/A	NO
26485243	C/G	NO
26485282	T/C	rs41276285
26485318	C/T	rs41276287
26485346	A/C	rs141241578
26490251	A/G	rs753508
26490251	A/G	rs753508
26490316	C/G	rs327217
26492397	C/T	rs78121726
26492470	G/T	rs55906521
26492529	C/T	rs76617667
26492560	A/C	rs1867042
26497684	A/G	rs191998425
26497820	A/G	rs77434938
26498013	T/G	rs13277175
26498023	G/A	NO
26499118	A/G	rs10088957
26499227	C/T	rs327218
26499331	C/T	rs118022714
26499497	C/T	rs62493399
26501169	*C/T*	rs187445527
26501179	G/A	rs187445527
26501344	A/G	rs184841156
26501351	A/G	rs7841384
26505135	G/C	NO
26505307	A/C	rs139163162
26506650	A/C	NO
26506809	C/T	NO
26510792	C/T	rs708621
26513312	T/G	NO
26513313	G/T	rs73227625
26513324	T/A	rs71216769
26513364	G/A	NO
26513525	G/A	rs76752934
26513708	T/A	rs35314544
26514054	G/C	rs113080437
26514101	C/T	rs58827210

26514144	C/T	rs17055639
26514171	G/A	rs17055641
26514231	C/G	rs17055646
26514293	G/A	rs1058332
26514316	C/T	rs11863
26515005	C/G	rs41276289
26515138	C/T	rs45471201
26515458	T/C	rs17666
26515472	G/A	rs10495
26515560	T/C	rs71656236
26515569	T/C	rs35258691
26515663	G/A	rs10042
26515695	G/A	NO
26515893	A/G	rs7827731
26516038	A/C	rs7828056
26516213	T/G	rs7831883
26518712	A/G	rs78578770
26518714	G/A	rs35346375
26518751	G/T	rs7845740
26518780	G/A	rs7842128
26518854	A/G	rs58047393
26518909	C/T	rs57846118
26518910	T/C	rs56322416
26520560	C/T	rs12155555
26520634	C/T	rs6557935
26535998	A/G	rs117181094
26536126	C/T	rs186111033
26536141	G/A	NO
26536143	G/A	rs7831201
26536337	G/T	rs138582766
26536442	C/T	rs73676401
26550225	C/G	NO
26550278	T/C	rs187288868
26550290	C/T	rs112917653
26550524	G/A	rs73678614
26550553	T/G	rs35568528
26550576	G/A	rs73678615
26550621	C/G	NO
26550818	T/C	rs73678616
26550866	T/C	rs73678617
26551345	C/G	NO
26551397	T/C	rs17330796

26551416	C/A	rs17055703
26551529	G/A	NO
26558088	C/T	rs190695344
26558113	T/C	rs17422356
26563796	A/G	rs17331399
26564046	T/C	rs193051886
26564049	T/C	rs17055758
26564127	G/T	NO
26564148	C/G	rs2044106
26564172	T/C	rs184835321
26564180	G/A	rs144125071
26564246	C/A	NO
26564320	T/C	NO
26564552	G/A	rs182164738
26564603	C/T	rs185508817
26564629	T/C	rs55780663
26564769	G/A	NO
26564949	A/G	rs56315822
26564953	A/G	rs71519610
26564966	G/T	NO
26565093	G/A	rs6985931
26565139	C/T	NO
26565140	G/A	rs116838475
26565155	C/T	NO
26565230	G/A	rs114428764
26565299	T/A	NO
26565429	C/T	rs116510299
26565510	C/T	rs115507492
26565666	A/G	rs191918596
26565686	A/T	rs7007717
26565770	C/T	rs7001655
26565771	C/A	rs115013410
26565845	T/C	rs111568746
26565846	A/T	rs71553819
26565854	A/G	rs34790461
26574356	C/A	rs79743265
26576602	C/T	NO
26577021	C/T	rs13278281
26577324	C/A	rs139839350
26577484	C/A	NO
26577645	C/T	rs73229635
26577795	A/T	NO

26577883	A/G	rs114104363
26577919	G/T	rs994098
26578017	G/A	rs1036813
26585092	G/T	NO
26585149	G/A	rs6998714
26585177	T/C	rs4732831
26585268	G/A	rs10503796
26585403	G/A	rs149880076
26585513	T/G	rs58777446
26585613	C/T	rs1908650
26585653	C/T	rs193100914
26586503	G/A	NO
26586689	A/G	NO
26586772	G/A	rs34099297
26586846	*A/G*	no
26586846	A/G	NO
26586846	A/G	NO
26586880	A/G	NO
26586931	C/T	rs1317256
26591848	C/T	rs1025860
26591865	G/A	rs186769997
26591883	C/A	rs1025859
26591911	A/T	rs141958536
26592057	T/C	NO
26592092	G/A	rs78167668
26592119	G/A	rs1025858
26592353	G/A	rs2132449
26592353	G/A	rs2132449
26592504	A/G	rs1016734
26592538	T/C	rs76802184
26592643	G/C	rs35551735
26592764	C/T	NO
26592765	G/A	rs190018049
26593159	C/T	rs17055838
26593179	T/C	rs111576315
26593200	C/T	rs191064950

Table 2.3: List of 68 rare and functional SNVs

SNV ID	Allele frequency (cases)	Estimated allele count (cases)	Allele frequency (controls)	Estimated allele count (controls)
26404986	0.015	11.52	0.015	11.52
26405344	0.003	1.92	0.000	0.00
26405508	0.066	50.56	0.046	35.20
26434583	0.005	3.84	0.004	3.20
26434626	0.014	10.88	0.005	3.84
26434888	0.000	0	0.003	1.92
26435213	*0.008*	6.4	*0.007*	5.12
26435238	0.000	0	0.002	1.28
26435451	0.006	4.48	0.003	1.92
26436000	0.001	0.64	0.000	0.00
26436034	0.067	51.2	0.053	40.96
26436221	0.002	1.28	0.002	1.28
26436344	0.002	1.28	0.000	0.00
26436349	0.016	12.16	0.013	9.60
26436351	0.003	2.56	0.000	0.00
26436783	0.062	47.36	0.053	40.32
26441264	0.013	9.6	0.016	12.16
26441529	0.003	1.92	0.000	0.00
26450631	0.003	1.92	0.000	0.00
26450858	0.056	42.88	0.035	26.88
26453162	0.005	3.84	0.003	1.92
26454793	0.062	47.36	0.058	44.80
26455539	0.000	0	0.047	35.84
26455543	0.000	0	0.004	3.20
26466011	0.000	0	0.013	9.60
26466520	0.003	1.92	0.000	0.00
26466789	0.001	0.64	0.000	0.00
26466803	0.003	1.92	0.000	0.00
26466877	0.072	55.04	0.067	51.20
26467541	0.025	19.2	0.023	17.92
26481666	0.002	1.28	0.000	0.00
26484061	0.010	7.68	0.006	4.48
26484081	0.002	1.28	0.000	0.00
26484146	0.001	0.64	0.000	0.00
26484392	0.000	0	0.003	1.92
26484404	0.034	26.24	0.023	17.28
26484426	0.027	20.48	0.043	32.64

26484652	0.002	1.28	0.003	1.92
26485346	0.003	2.56	0.000	0.00
26490251	0.002	1.64	0.002	1.64
26498023	0.003	1.92	0.002	1.28
26499497	0.057	43.52	0.034	26.24
26501344	0.005	3.84	0.000	0.00
26505307	0.003	1.92	0.002	1.28
26513324	0.032	24.32	0.063	48.00
26513364	0.013	9.6	0.013	10.24
26513525	0.008	6.4	0.026	19.84
26514054	0.016	12.16	0.028	21.76
26514101	0.016	12.16	0.017	12.80
26514144	0.017	12.8	0.016	12.16
26514171	0.017	12.8	0.017	12.80
26514231	0.067	51.2	0.054	41.60
26515695	0.003	2.56	0.002	1.28
26551529	0.003	1.92	0.000	0.00
26564603	0.013	9.6	0.004	3.20
26565429	0.000	0	0.003	1.92
26565510	0.000	0	0.003	2.56
26574356	0.051	39.04	0.042	32.00
26576602	0.000	0	0.001	0.64
26577324	0.022	16.64	0.021	16.00
26577484	0.001	0.64	0.000	0.00
26577795	0.000	0	0.001	0.64
26577883	0.076	58.24	0.044	33.92
26591911	0.018	14.08	0.033	24.96
26592643	0.053	40.96	0.034	26.24
26592764	0.016	12.16	0.016	12.16
26592765	0.008	6.4	0.008	6.40
26593200	0.029	22.4	0.017	12.80

Table 2.4: Variant enrichment analysis

Allele Frequency	Fold-Change	p-value
1%	1.28	5×10^{-4}
2%	1.07	0.18
5%	1.05	0.15
10%	1.08	4×10^{-4}

Table 2.4: After filtering the total SNVs for rare and functional variants, we analyzed 68 variants for enrichment in cases vs. controls. We identified a modest enrichment for rare variants in SZ cases at 1% and 10% allele frequencies ($p<4x10^{-4}$).

Figures: Chapter 2

Figure 2.1: Strategy for DPYSL2 sequencing screen

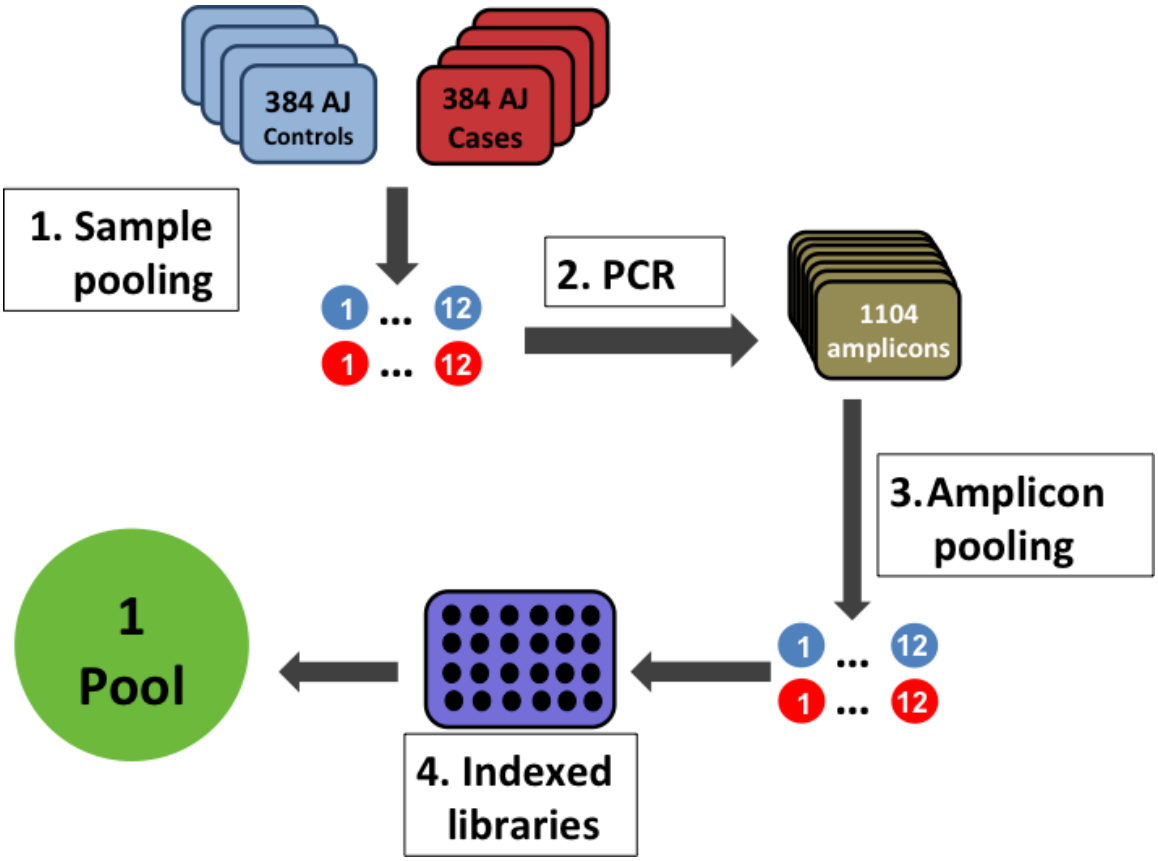

Figure 2.1: Strategy used for *DPYSL2* sequencing screen. **(1)** AJ samples were normalized amd mixed into 12 pools of cases and 12 pools of controls, each with 32 samples. **(2)** For each pool I did 46 PCRs to span the 14 exons and 27 conserved non coding regions of *DPYSL2*. **(3)** The amplicons from the same pool were combined into 24 pools, each with 46 amplicons. **(4)** The DNA from 24 pools were randomly fragmented, and used to generate 24 indexed libraries with Illumina's Truseq kit. Indexed library were quantified and combined together into 1 pool, which was

then sequenced on one lane on the Genome Analyzer to get 75 bp paired end reads.

Figure 2.2: Distribution of sequenced reads

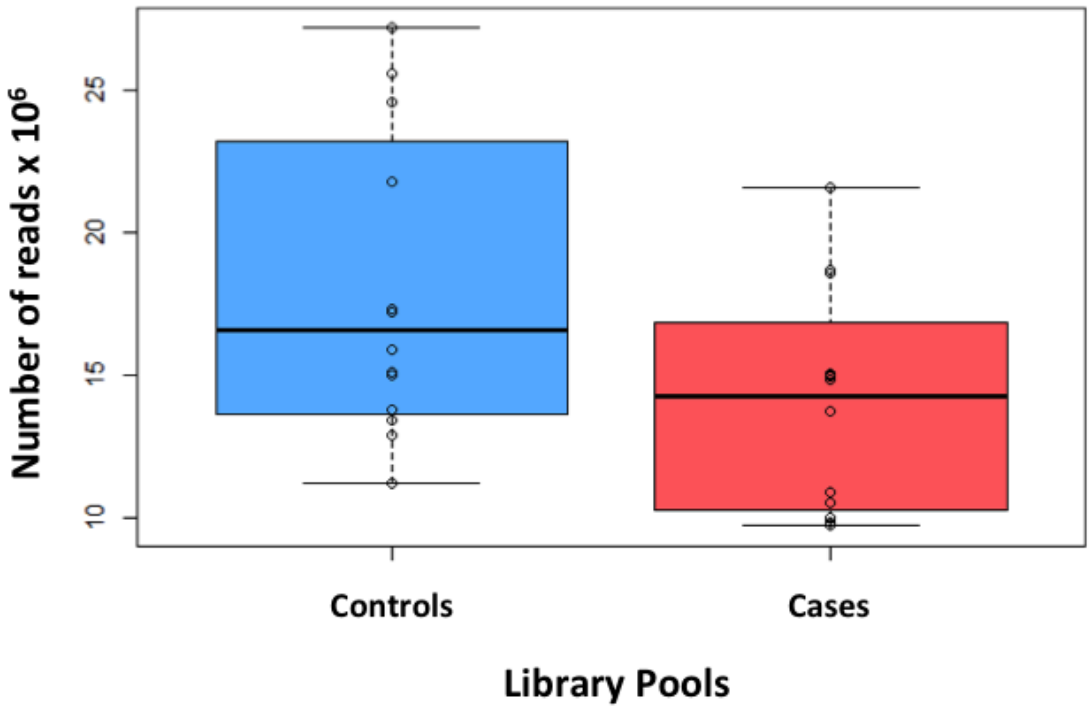

Figure 2.2: the distribution of reads for all 24 pools, separated by case and control pools. The differences in distribution between cases and controls are not significant, and the average number of reads per pool was 16.3 million.

Figure 2.3: Distribution of identified variants

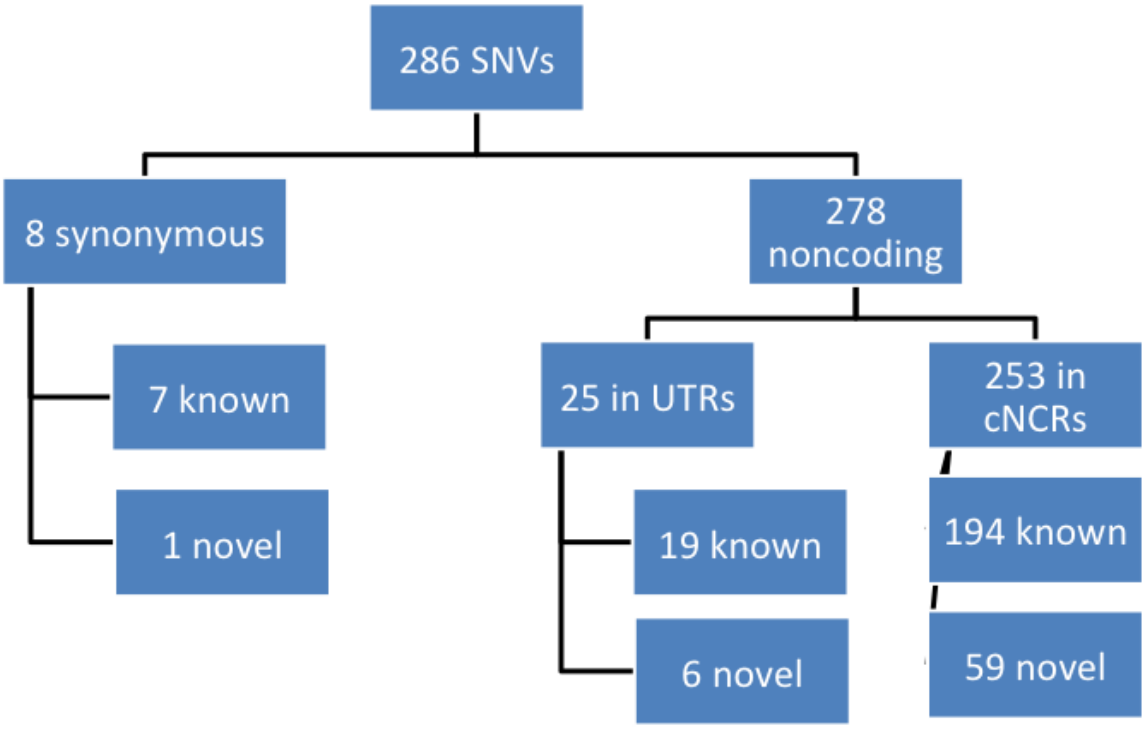

Figure 2.3: A total of 286 SNVs were identified 286 using the manual and automated methods of variant identification.

Figure 2.4: Splicing analysis of 2 synonymous variants

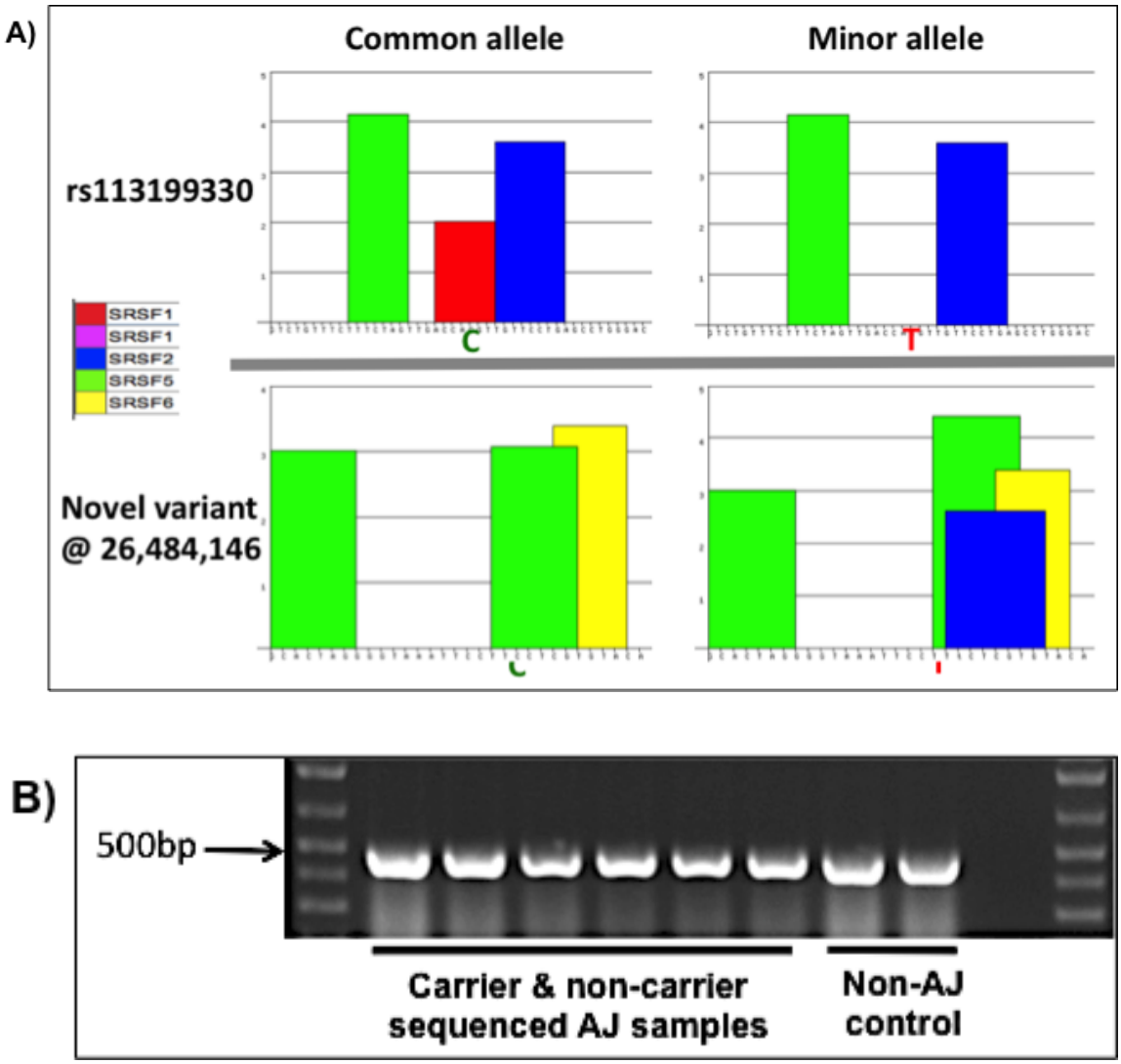

Figure 2.4: (A) Synonymous variants were predicted to alter binding of splicing factors. **(B)** PCR of carrier and non-carrier DNA obtained from lymphoblasts do not show altered splicing patterns.

Chapter 3: The *DPYSL2* gene connects mTOR and schizophrenia

Xuan Pham [1,2], Guang Song [3], Serena Lao [4], Loyal Goff [2], Heng Zhu [3], David Valle [1, 2], Dimitrios Avramopoulos [1, 2]

1) McKusick-Nathans Institute of Genetic Medicine; 2) Predoctoral Training Program in Human Genetics; 3) Department of Pharmacology and Molecular Sciences; 4) University of Maryland, College Park; Johns Hopkins University School of Medicine, Baltimore, MD 21205

Abstract

We previously reported a schizophrenia-associated polymorphic CT di-nucleotide repeat (DNR) at the 5'-UTR of *DPYSL2* which responds to mTOR signaling with allelic differences. Using microarray analysis, we now show the DNR alleles interact differentially with specific proteins, including the mTOR-related protein HuD/ELAVL4. We confirm the differential binding to HuD and other known mTOR effectors by electrophoretic mobility shift assays. In HEK293 cells edited by CRISPR/Cas9 to carry the schizophrenia risk variant (13DNR) we observe significant reduction of the corresponding CRMP2 isoform and shortening of the cellular projections. Additionally, transcriptome

analysis by RNA-seq shows changes in 12.7% of expressed transcripts. These transcripts are enriched in immunity related genes, overlap significantly with those modified by the schizophrenia-associated gene, *ZNF804A,* and have a reverse expression signature from that seen with antipsychotic drugs. Our results support the functional importance of the *DPYSL2* DNR and a role for mTOR signaling in schizophrenia.

Introduction

Family, twin, and adoption studies indicate a strong genetic contribution to risk for SZ [5,6], and in a linkage analysis of 57 multiplex pedigrees covering ~70% of the genome we previously identified region 8p21-22 as a candidate SZ susceptibility locus (p=0.0001) [9]. Our result was replicated by many other groups in different populations [10–12,36], and we followed it up with a targeted association study in an Ashkenazi Jewish population [13]. We found the most significant SNP (rs12155555) was located ~5 kb upstream of *DPYSL2B,* and the strongest association signals from the fine mapping were located within or flanking the *DPYSL2* gene [13].

Subsequently, we identified multiple functional sequence variants in *DPYSL2B,* including a polymorphic CT di-nucleotide repeat (DNR) variant located within the gene's 5'-TOP tract that's known to be regulated by the mTOR pathway [30,31,40,41]. In Caucasians, this DNR most often comprises 11 CT repeats (11DNR). The next most common allele caries 13 CT repeats (13DNR) and previously, in a small sequencing study, we found it at a frequency of ~14.5% in 93 healthy Caucasians and 29% in 46 Caucasians with SZ [42].

The mTOR complex 1 (mTORC1) is a serine/threonine kinase complex that controls protein synthesis by phosphorylating downstream target effectors (Fig. 3.1). The complex owes its name to the potent

immunosuppressant drug, Rapamycin, which strongly inhibits mTORC1 kinase activity. Canonical activation of mTORC1 results from an activation cascade of upstream proteins, including receptor tyrosine kinases (RTK), PI3K, and AKT. Once active, mTORC1 stimulates its downstream effector proteins, S6 and 4EBP, increasing ribosome biogenesis and mRNA translation, respectively [43–45]. Fundamentally, mTOR activation boosts overall cell growth and metabolism, marking it as one of the better-known targets for cancer therapy [46–49]. Increasingly however, the functions of mTOR have been extended to the development of the central nervous system, where mTOR is implicated in neuronal growth, maintenance, and proliferation [50–53]. Perturbation of the mTOR signaling cascade can affect well known neurotransmitters involved in schizophrenia, such as serotonin, glutamate, and their receptors. Thus, disruption or dysfunction of this pathway could play a role in the risk to develop SZ [51].

In this chapter, I report evidence that extend and strengthen the functionality of the risk 13DNR variant and its regulation by mTOR. First, in dual luciferase and polysome profiling assays we showed that the 13DNR variant significantly reduced expression at the translation level in HEK293 cells and in mouse primary cortical neurons. Further, in both cell types, there was a dose-dependent response of both DNR alleles to Rapamycin, an allosteric inhibitor mTOR [30]. Then, we characterized the 5'UTR of the *DPYSL2* transcript and its variants for interaction with

specific proteins, including mTOR effectors. Next, I showed that this naturally occurring DNR variant also has strong effects on *DPYSL2* regulation in its native context in the genome, with the longer repeat leading to significant reduction of the corresponding CRMP2 isoform in HEK293 cells. This change is accompanied by striking shortening of the natural HEK293 projections. Finally, we showed that significant changes in the transcriptome point to pathways implicated in SZ; these overlap with the changes observed with other SZ susceptibility genes, and oppose those observed by exposure to antipsychotics.

Materials and Methods

Cell culture

HEK293: The human embryonic kidney cell line (HEK293; ATCC #CRL-1573) was cultured using standard protocols. Cells were maintained in Dulbecco's Modified Eagle Medium (Invitrogen #11965-084) with 10% fetal bovine serum (Sigma #12003C).

Mouse primary cortical neurons: I isolated and cultured mouse primary cortical neurons from E18 C57BL/6 mice using a previously described protocol [54]. Cells were maintained in Neurobasal medium (Invitrogen # 21103049), 5% B-27 supplement (Thermo Fisher #17504044), 1% Penicillin-Streptomycin antibiotic (Thermo Fisher

#15140122), and 200mM L-glutamine (Thermo Fisher #25030081). Cells were grown on coverslips coated with poly-D lysine (Sigma #P0899) in borate buffer.

Dual luciferase assays

To test the mTOR regulation in the 5'-UTR of DPYSL2, we transfected reporter constructs into HEK293 and mouse primary cortical neurons in the presence of increasing concentrations of Rapamycin. The reporter constructsfor the 11 and 13DNR alleles were the same ones generated previously [30]. These were transfected into cell using Lipofectamine (Thermo Fisher #11668030) for 4 hours, after which the culture media was replaced to media containing the indicated concentrations of Rapamycin (Selleck Chem #S1039), dissolved in DMSO. After 24 hour exposure to the drug, we analyzed the cell lysates for luciferase activity.

Protein interaction study

Protein microarray: To study the differential binding of proteins to the DNR alleles, we performed protein microarrays. Fluorescently labeled RNA probes for the 11 and 13 DNRs were synthesized by IDT (Table 3.1). The probes were hybridized to the protein microarrays in duplicates and as described in previous protocol [55]. The microarray chips contained >4,000 proteins, each of which is printed in duplicates, and

with multiple sets of positive control proteins. After hybridization, the protein chips were scanned with a GenePix 4000B scanner with 5-μm resolution detection at 635 nm. Microarray images then underwent corrections and normalizations, as specified in a previous protocol [56] . We selected an intensity cutoff value of Z-score > 9 to produce a significantly reduced list of "positive hits." (Fig. 3.2) Proteins were considered to bind both alleles if the difference between the two allelic averages was less than two standard deviations. If the difference was greater than two standard deviations, then the protein was considered to bind preferentially to either the 11 or 13 DNR allele.

EMSA: In order to confirm the differential binding of ELAVL4 and to test the binding of other mTOR signaling components such as S6K, 4EBP and eIF4E, we performed electrophoretic mobility shift assays (EMSA). I utilized the same 3' fluorescently labeled RNA probes from the microarray, but end-labeled the probes with radioactive γ–ATP-32 to the 5' end [57]. The probes were purified (VWR SpinX Filters #29442-754, Clonetech Resin #631770), and precipitated overnight at -80C (3ul yeast tRNA, 6ul, NH_4Ac, 24ul cold 100% EtOH). Radioactivity of the probe was quantified with a scintillation counter, and each probe was normalized to 4,000 counts per minute (CPM).

The radioactive RNA probes for each of the two DNR alleles were hybridized to commercially purified candidate binding proteins: S6

(Origene #TP317324), 4EBP (NovoProtein #C200), eIF4E (MyBioSource #MBS203380), and ELAVL4 (Origene #TP318612). Each probe was incubated on ice for 20 minutes in the following solution: 2X Probe Buffer (50% glycerol, 1M HEPES, 1M Tris-HCL, 0.5M EDTA, 100mM DTT, DEPC H2O, Bromophenol blue), BSA (10mg/ml, NEB #B9000s), Poly[d(I-C)] (1ug/mL, Roche #10108812001), DEPC H20, γ–ATP RNA probes (1ul at 4000 CPM), and purified proteins (1ug/ul). Each reaction was loaded into a precast 5% TBE gel (Bio-Rad #4565015) and electrophresed in 1x TBE buffer (Bio-Rad #1610733) at 150V for 30 minutes, room temperature. The gel was air dried in clear cellophane (Bio-Rad #1651779) for 3 hours, and then exposed with autoradiography film (VWR # 8294985).

CRISPR/Cas9 Targeting

To study the effect of the DNR in its genomic context, we used the CRISPR/Cas9 system to create two isogenic cell lines with the 11 and 13 DNR in HEK293 cells, which normally carry the 11 repeat DNR. Using the online tool provided by the Zhang Lab at MIT [58] (crispr.mit.edu), we designed three guide RNAs (gRNA) (Table 3.1). As a repair template to insert 2 CT repeats at the DNR region, we used the 13DNR reporter constructs from the dual-luciferase assay [30]. The gRNA constructs were synthesized with the U6 Promoter and gRNA backbone via the gBlock® system by IDT [59].

We used lipfectamine (Thermo #11668027) to co-transfect the gRNA, the repair template, and a GFP tagged Cas9 nuclease (Addgene #44719) to HEK293 cells. Post transfection, the cells underwent fluorescence-assisted cell sorting into 96-well plates (1 cell/well). The single-cells were then expanded into clonal colonies, the DNA of which was used for sequencing to identify homozygotes for each allele (Fig. 3.6).

We then selected 17 gene loci as candidates for off-target effects predicted by the online tool developed by the Zhang Lab at MIT to check for off-target effects. Using primers (Table 3.1) designed to capture and Sanger sequence 100-200 bp of these candidate loci, we identified no known off-target effects of the gene-editing scheme.

Western blot

We obtained cell lysates from the four 13DNR homozygote clones and four targeted 11DNR control clones, using 1X Passive Lysis Buffer (Promega #E1941). Protein was assayed with a colorimetric kit (Bio-Rad #5000001), quantified with spectrophotometry, and lysates were normalized to 20ug total protein.

Protein was heated to 95C for 5 minutes in the following solution: 2x loading buffer (4% SDS, 10% 2-mercaptoethanol, 20% glycerol, 0.004% bromophenol blue, 0.125 M Tris-HCl), H_20. The samples were loaded into precast gels (Bio-Rad #4561083) and electrophresed at 120V

for 1 hour in 1x running buffer (Bio-Rad #1610732). The gel was then transferred to a PVDF membrane (Bio-Rad #1620177) via the semi-dryelectrophoretic transfer cell (Bio-Rad #1703940) at 15V for 45 minutes.

After transfer, the membrane was blocked in 5% powder milk and 1x TBS buffer (Bio-Rad #1706435) for >1 hour at room temperature. We then probed the membrane using rabbit anti-CRMP2 antibody at 1:5000 (Sigma #C2993) and mouse anti-GAPDH at 1:4,000 (AbCam #ab9485) in 1% milk overnight at 4C. Following wash steps in 1x TBS-T, we probed with mouse (AbCam #ab6728) and rabbit (Thermo #656120) HRP-conjugated secondary antibodies at 1:10,000 in 1% milk for >2 hours at room temperature. After final washes, the membranes were exposed to chemiluminescence for HRP detection (Thermo #32132), and then exposed to autoradiography film.

For Rapamycin exposures, we cultured the edited HEK293 cell in standard protocols with media containing 30nM Rapamycin (SelleckChem #S1039) for 24 hours. After drug exposure, we obtained the cell lysates as described above.

Real-Time qPCR

RNA was obtained from the selected CRISPR-edited clones using an RNeasy Mini Kit (Qiagen #74104). The RNA was reverse transcribed

into cDNA using qScript cDNA Supermix (Quanta Biosciences #84033). From this, we performed qPCR using PerfeCTa SYBR Green FastMix (Quanta Biosciences #95072) and primers for B-actin and DPYSL2. Samples were run on the 7900HT Sequence Detection System by Applied Biosystems.

Immunofluorescence & Imaging

Targeted HEK293 cells carrying the homozygous forms of the 11 and 13 DNR alleles were grown on Poly-D Lysine (Sigma #P6407) coated coverslips in 12-well plates in standard HEK293 media at 37C. The media was changed every other day until day four (~75-85% confluency). Cells were washed twice with 1x Passive Buffered Saline, and then fixed in 4% paraformaldehyde for 15 minutes at room temperature. After removal of the paraformaldehyde and washing, the cells were incubated with 1mg/ml wheat germ agglutinin conjugated with Alexa Fluor 488 (Thermo #W11261) at room temperature for 15 minutes. After removal and washing, the coverslips were mounted on slides with mounting media containing DAPI stain (Vectorlabs #H1200).

Eighteen slides (nine from each cell group) were randomized and blinded before imaging. We had 4 clones per genotype and 3 slides per clone. We collected eight images for each slide. The images were analyzed blinded for cellular projection length differences using ImageJ.

RNA-seq

We obtained cell pellets for the four 13DNR homozygous edited clones, and eight targeted 11DNR control clones. RNA was extracted from the cell pellets using an RNeasy Mini Kit (Qiagen #74104). The RNA was quantified and normalized to 500ng per sample and the samples were then sent to the Hopkins Core Facility. Agilent BioAnalyzer was used for quality control of the RNA prior to library creation, with a minimum RIN of 8.5. RiboMinus™ technology (ThermoFisher Scientific) was used to deplete ribosomal RNA molecules and Illumina's TruSeq RNA v2 protocol was used to generate libraries. Specifically, total RNA was converted to cDNA and size selected to 150 to 200 bp in length with 3' or 5' overhangs. End repair was performed where 3' to 5' exonuclease activity of enzymes removes 3' overhangs and the polymerase activity fills in the 5' overhangs. An 'A' base was then added to the 3' end of the blunt phosphorylated DNA fragments which prepares the DNA fragments for ligation to the sequencing adapters and barcodes, which have a single 'T' base overhang at their 3' end. Ligated fragments were subsequently size selected through purification using SPRI beads and undergo PCR amplification techniques to prepare the 'libraries'. The BioAnalyzer is used for quality control of the libraries to ensure adequate concentration and appropriate fragment size. The resulting library insert size was 120-200bp with a median size of 150bp. DNA sequencing was performed on an Illumina® HiSeq 2500 instrument using standard protocols for paired

end 100bp sequencing. As per Illumina's recommendation, 5% PhiX was added to each lane as a control, and to assist the analysis software with any library diversity issues. Primary Analysis was also performed in our core facility. Reads were processed through Illumina's Real-Time Analysis (RTA) software generating base calls and corresponding base call quality scores. CIDRSeqSuite 7.1.0 was used to convert compressed bcl files into compressed fastq files.

RNA-seq data analysis

Forward and reverse raw reads separated by 12 barcode were received from our core facility showing no problems in data quality according to the accompanying FastQC reports (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Analysis was performed using the Tuxedo suite of programs including tophat for read mapping, cufflinks for transcript assembly and cuffdiff for transcript abundance comparisons as described previously [60]. Following the Tuxedo pipeline the output was further processed using the R Bioconductor package CummeRbund (http://bioconductor.org) for analysis and graphical visualization.

Results

The DPYSL2 5'TOP sequence responds to mTOR

The polymorphic DNR, located a 5'-TOP region, was previously shown to reduce *DPYSL2* expression at the translational level. Because translation of 5'-TOP mRNAs is regulated by the mTOR pathway, we sought to explore how regulation of *DPYSL2* expression may be influenced by mTOR signaling. We performed dual luciferase assays with HEK293 transfected the 11DNR and 13DNR in the absence and presence of Rapamycin, a potent mTOR inhibitor. In the absence of inhibitors, our results recapitulated previous data with the Risk allele showing ~2-3 -fold decrease in luciferase activity as compared with the WT allele ($p<1.3 \times 10^{-8}$). At concentrations of Rapamycin ranging from 0-30 nM, the 13DNR allele showed lower luciferase expressions, as compared to the 11DNR allele. Concentrations higher than 30nM substantially reduced luciferase expression for both alleles to a similar extent (Fig. 3.3A). The same trend was observed for mouse primary cortical neurons (Fig. 3.3B) From this we concluded that the differences in *DPYSL2* expression between the two DNR alleles are likely mediated by differential binding of mTOR effectors [30].

The DNR alleles differentially bind TORC1 effectors and HuD

To identify proteins binding to the 5'TOP sequence of *DPYSL2*, we used protein microarrays we have previously developed [55]. We found exclusive binding of the 11DNR to 5 proteins: CHAMP1, a chromosome alignment phosphoprotein necessary for accurate mitotic division, and whose mutations have been recently linked to syndromic intellectual disability [61,62]; CAST, a calpain inhibitor involved in proteolysis, and whose mutations have been implicated in Alzheimer's disease [63,64]; LMOD3, a member of the leiomodin family involved in skeletal muscle organization [65]; TCEAL5, an X-linked transcription elongation factor [66]; and ELAVL4 a well characterized ribosomal binding protein (RBP) with exclusive neuronal expression [67] (Table 3.2). *ELAVL4*, also known as *Human Antigen D (HuD)*, is a member of a four-gene family of RBPs. Its known functions include the transport of target mRNAs from the nucleus to the cytoplasm and the triggering of neuronal differentiation [67–69]. Most interestingly, recent evidence shows that like *DPYSL2B*, *ELAVL4* is regulated by the mTOR pathway and promotes dendritic branching [70].

Electrophoretic mobility shift assays (EMSA) with nuclear and whole cytosolic HEK293 cell extracts showed clear shifts for both oligos, indicating that the DNR probes have the capacity to bind to whole proteins (Fig. 3.4A). Then, in the presence of purified specific proteins,

the audiograph confirmed the differential binding of the 11DNR to ELAVL4 and to two additional proteins that are part of the mTOR downstream pathway, 4EBP and eIF4E (Fig. 3.4B). We noted the presence of an additional band migrating slightly higher than the main, unbound oligo band for the 11DNR allele. In subsequent additional experiments, we found that this band could be significantly diminished with heat treatment (Fig. 3.5A) immediately before loading (unfortunately this step cannot be incorporated in the EMSA experiment), suggesting potential dimers or folding. In later experiments using new batches of both RNA oligonucleotides, we observed the same bands for both 11DNR and 13DNR, and both could also be resolved with heat (Fig. 3.5B). We used these new oligos in additional EMSA experiments, repeating the initial and adding protein combinations (Fig. 3.5C). In those experiments the additional bands were present for both alleles, yet there was no difference from our original results. Given these experiments, we have excluded that this likely dimer formation or folding does not produce artifacts that could be responsible for the differences we see in binding.

The 13DNR allele expresses reduced CRMP2 protein

To test whether the differences we observed between the DNR alleles in reporter assays also apply to the translation of the *DPYSL2B*B transcript itself, we introduced the 13DNR allele into HEK293 cells using CRISPR/Cas9 genome editing technology [59]. After targeting, we obtained

112 HEK293 cell clones that were successfully transfected and flow sorted for the presence of the co-transfected GFP tag. Of these, 4 clones (3.6%) were 13DNR homozygotes, and 3 clones were (2.7%) were heterozygotes (Fig. 3.6). The remainder clones were not modified at the target site, and were used as 11DNR controls. We selected 8 of these controls and the 4 homozygous 13DNR clones for further sequencing at 17 candidate off-target gene sites, and identified no off-target mutations.

Western blot analysis on 4 DNR11 and 4 DNR13 clones using a CRMP2 antibody recognizing all transcripts showed significant differences between the two isogenic cell groups (Fig. 3.7A, C). CRMP2B, migrating at 64 kDa, was significantly reduced in the 13DNR cells as compared to the 11DNR cells, both relative to CRMP2A migrating at 72 kDa ($p=6.5x10^{-6}$) and in absolute terms after normalization to GAPDH ($p=0.0077$) (Fig. 3.7A, C). The absolute levels of CRMP2A were not affected ($p=0.826$).

After exposure of the cells to 30nM Rapamycin, levels of CRMP2B protein in the 13DNR clones were reduced by nearly 8-fold ($p=2.8x10^{-5}$) compared to only a 1.6-fold change for the 11DNR clones ($p=0.0014$) (Fig. 3.7B, C). The reduction recapitulates our previous observations in applying Rapamycin to *DPYSL2* promoter constructs driving luciferase expression [30]. Overall, the results suggest that the 13DNR risk allele has

a weaker response to mTOR signaling, and is more sensitive to mTOR inhibition.

In contrast to the protein results, the transcript profiles for *DPYSL2* were not significantly different when measured by *DPYSL2* specific RT-PCR between the 11 and 13DNR cells, suggesting that the DNR affects primarily translation as expected for a variant in the 5'-TOP sequence in response to mTOR signaling.

13DNR results in shorter cellular projections

HEK cells naturally show neurite-like projections and have been suggested to have a neuronal lineage due to the original transformation methodology [71]. They have also been found to express multiple neuron specific markers [71]. Following the introduction of the 13DNR, we observed that the HEK293 projections appeared shorter in the 13DNR homozygous cells (Fig. 3.8A). This observation was confirmed by blinded measurements of projection length and comparison between homozygous 11DNR cells (n=283) to homozygous 13DNR cells (n=348, $p=10^{-42}$) using imageJ [72]. The estimated length difference was ~2-fold (Fig. 3.8B).

13DNR alters transcription, differences point to immunity genes

The cell's transcriptional output is the first level of response to any change, be it genetic or environmental. Having shown that the DNR variant has a strong effect on isoform abundance, we then asked whether

this change is important for the cell's homeostasis as reflected in its transcriptome, and whether the observed changes support a role in disease. We performed RNA-seq of the two isogenic cell groups acquiring ~20 million paired reads for each of the 12 individually targeted clones (four with 13DNR and eight with 11DNR). The reads showed 85-90% unique alignment to the human genome. Initial quality control showed that all samples had similar fragments per kilobase per million (FPKM) distributions, but the 11DNR samples showed significantly higher variance across all FPKM levels (Fig. 3.9A). Additionally, a dendrogram based on all genes showed that two 11DNR samples (C2_5, C2_6) clustered together and separated early from all other samples, suggesting they are outliers (Fig. 3.9B). We considered that this may be either due to undetected off-target editing or to the chromosomal instability of this neoplastic cell line. Removing them restored the variance of the remaining six 11DNR samples to levels similar with the 13DNR. A new dendrogram using all genes on the new set of samples clustered the 11DNR and 13DNR samples separately, showing that the genome wide expression profile differences were driven mostly by the genotype difference at the *DPYSL2B* DNR (Fig. 3.9C, D).

From a total of 13,548 genes that were expressed in these cells, 1,730 (12.7%) showed expression differences (FDR <0.05) between the 11DNR and 13DNR groups. Of those, 802 were higher in 13DNR and 928 in 11DNR. We used Panther bioinformatics to test whether the up- and

down-regulated genes showed enrichment for membership to different functional categories compared to the whole set of expressed genes. The most significant enrichment in biological processes was for genes expressed lower in the 13DNR that were often immune system process genes (2.04 fold, $p = 2.6x10^{-7}$). This mirrors enrichments seen in SZ genes by both GWAS [73] and proteomics analyses [74–76]. Genes expressing higher in 13DNR were strongly enriched for those encoding RNA binding proteins (2.6 fold, $p=5.25x10^{-10}$). The complete results at $p<0.01$ are shown in Table 3.3.

Genes changed by 13DNR overlap with genes changed by ZNF804A knockdown

A previous study by Chen et al. [77] used a similar approach to ours on another SZ associated gene, *ZNF804A*. That study performed RNA-seq analysis after a knockdown approach to reduce the expression of, *ZNF804A* in neural progenitor cells derived from induced pluripotent stem cells. Reduced expression of *ZNF804A* has been reported to be the consequence of the SZ-associated allele of SNP rs1344706 in this gene [78]. Despite significant differences between the studies, when we compared the reported up- or down-regulated genes by that study with ours, we found highly significant concordance and overlap. Specifically, of 57 genes whose expression were significantly altered, 45 were in the same direction for both studies (binomial test $p = 6.5x10^{-6}$). The overlap was significantly more than expected by chance for genes that were expressed

higher in the SZ risk condition, with 37 genes in common, more than twice more from the expected 16 (Fisher's exact test $p= 1.6x10^{-6}$), but not in those expressed lower.

13DNR transcription signatures are opposite to antipsychotic drugs

The Connectivity Map (cmap) [79,80] is a collection of genome-wide expression data from human cell lines treated at varying concentrations in culture with 1,309 different bioactive small molecules, 18 of which are antipsychotics. Pattern-matching computational tools integrated in the cmap website (www.broadinstitute.org/cmap) allow the user to upload any gene signature (a set of up- and down-regulated genes) and compare it with the signatures of these small molecules for positive or negative correlations. We did this analysis for the transcriptional signature of the 11DNR/13DNR variant and found that 3 of the top 10 cmap signatures most significantly matching the 13DNR group (all $p < 2 x10^{-5}$), 3 were the antipsychotics Thioridazine, Trifluoperazine and Prochlorperazine. All 3 had complimentary (i.e. opposite) signatures to the 13DNR allele. Interestingly, Rapamycin, the mTOR inhibitor, was also one of these top 10 signatures and showed complimentary signature to the 13DNR risk allele.

Discussion

We demonstrate here that the DNR variant in the 5'UTR of the *DPYSL2B* transcript has strong effects on the expression of the corresponding protein isoform, and provide additional evidence that this is likely due to changes in its response to mTOR signaling. We find reduced binding to the 13DNR risk allele of two downstream mTOR effectors, 4E-BP and eIF4E, which may mediate these changes. In eukaryotes, phosphorylated 4E-BP releases eIF4E, allowing the de-repressed factor to initiate cap-dependent translation [47,81]. We also find that the binding of ELAVL4 to the *DPYSL2* 5'TOP alleles is markedly different comparing the 11DNR to the 13DNR alleles. The striking similarities in the functions of CRMP2 and ELAVL4 in neurons and their common regulation by mTOR make this an interesting lead for further research.

The mTOR pathway promotes cell growth and division, and is widely known to be involved in cancer progression [46,82]. Mounting evidence indicates that the mTOR pathway is also involved in brain plasticity, with aberrant signaling contributing to several neurodevelopmental disorders including schizophrenia [51–53]. Our previous and current results support this hypothesis. We show that *DPYSL2*, a candidate schizophrenia susceptibility gene that's involved in axonal growth and neuronal development, is regulated by mTOR. The

response to Rapamycin, the binding of mTOR effectors, and the transcriptome changes in the presence of the weaker binding allele all demonstrate the regulation of *DPYSL2* translation by mTOR. The downstream effects of these changes provide candidate mechanisms for the connection to neurodevelopment and SZ.

The link between mTOR and SZ risk is particularly interesting because it provides a potential explanation for environmental disease risk factors. The importance of starvation in SZ risk has been strongly established by studies in the Dutch and Chinese populations that experienced severe famines in the early to mid 20th century [83,84]. In these populations, prenatal famines led to a 2-fold increased risk for schizophrenia. The mTOR pathway is highly sensitive to external environmental factors and, in particular, the availability of nutrients [85,86]. Starvation or otherwise malnutrition negatively affects the activity of this pathway, and could conceivably increase schizophrenia risk through changes in the regulation of genes involved in nervous system development such as *DPYSL2*.

We used genome editing by CRISPR/Cas9, and created two isogenic cell lines differing genetically only in the DNR (CT) repeat length at the 5'UTR of *DPYSL2*. We observed that homozygotes for the 13DNR variant produced significantly less CRMP2B protein. Consistent with the known functions of CRMP2B, they also had significantly shorter cellular

projections. The transcriptome differences between 11 and 13DNR cells show that many genes are affected by this genetic change, presumably as a result of the change in CRMP2B production. These genes, listed in Supplementary table 1, provide important clues for unraveling the pathways involved in the function of CRMP2 and its possible connections to disease.

In previous work [42], we found in a small sample that the 13DNR allele frequency was twice as high in SZ patients as in controls. To get a better estimate of the effect in a large sample we have identified two SNPs (rs367948, rs445678) that are in imperfect LD with this repeat, predicting the repeat genotypes in 41 of 42 individuals with data (data not shown). In the results of the PGC2 GWAS [73] these show a small increase in risk of 1.1 fold (p=0.04 - p=0.009 respectively in the same direction). It is remarkable that a variant with such profound effects on the transcriptome and on cell morphology in cultured cells has such a small effect on the phenotype of the complete organism. This observation suggests that strong buffering mechanisms in the organism are involved to minimize the consequences of the DNR variant. This is not only good news for the carriers of such risk variants, but also for researchers exploring them, as it suggests that despite the small effect on risk typical of most GWAS-identified variants, their effects on isolated cellular models may be strong enough to measure and interpret.

In our transcriptome analyses we found significant changes in 12.7% of transcripts resulting from the 13DNR. The specific genes affected by this variant provided three new lines of support for a connection between the gene and SZ. First, among down-regulated genes, the strongest enrichment was for genes involved in immunity. Although not directly intuitive, the immune system has been repeatedly implicated in SZ including the largest GWAS to date [73], proteomic studies [23,87], gene expression studies [88], and reports for a role of infection/inflammation [89,90]. Second, we saw a strong overlap and directional consistency between the genes affected by the SZ risk allele in our study and those affected by a similar perturbation in a previously published study of another SZ gene, *ZNF804A*. As more genes are studied, such overlaps should highlight the most important connections between risk variants and the corresponding pathways.

Finally, in a survey of the 1,309 pharmaceutical compound-induced expression change signatures in the connectivity map database, three antipsychotics raised to the top significance, all with signatures opposite to that of the 13DNR risk allele. This not only supports our hypothesis connecting the DNR to SZ but most importantly it indicates how cellular models might be useful in drug discovery. Other compounds on the top of the list would be interesting for further exploration of potential benefits against psychosis.

In conclusion, our experiments show that the SZ candidate gene transcript *DPYSL2* is regulated by the mTOR pathway and this regulation is perturbed by a 5'UTR SZ-associated DNR, adding support for mTOR's role in psychiatric disease. Further, our study of the transcriptome of genome-edited cells not only confirms the biological significance of the DNR, but it also provides additional links between *DPYSL2* and schizophrenia genes, pathways, and treatments.

Tables: Chapter 3

Table 3.1: CRISPR/cas9 oligos

gRNA sequences		Product size (bp)
DNR_gRNA1	TCTTTTTTTTCCGCCCTAGC	--
DNR_gRNA2	CTTTTTTTTCCGCCCTAGCT	--
DNR_gRNA3	TTTTTTTTCCGCCCTAGCTG	--
Genotyping primers		
HekDNR_250_F	TGCCTGAGAGGAAAGGGAGT	250
HekDNR_250_R	AATAGCAAGACCAGCGAAGC	250
Off-Target sequencing primers		
OTGprimer1F	TACACCTCCGTTGCCCAGgc	288
OTGprimer1R	AGTCTGGCAGGAGTGATGAC	
OTGprimer2F	CAGTTCCTTAGCAATGCACTGA	278
OTGprimer2R	CATGGGAATTGGCTGAGACC	
OTGprimer3F	TCCCTCATCTTTGAATGTGTGC	260
OTGprimer3R	CCCATGATGCACCAGAAAGAG	
OTGprimer4F	AGTCCCCGCCTTCGCTAGgg	287
OTGprimer4R	CAAGGGCTTTTGAGTTTATCGAG	
OTGprimer5F	CGCTTCTGTAGGGCCTTCAT	285
OTGprimer5R	CCTTCTTGTCGCAGTGGAAC	
OTGprimer6F	TTCTCCAAGGAAACCAACATGT	263
OTGprimer6R	TATCCATTTGACCAGCCCCT	
OTGprimer7F	ATTCATCCGCGACATCCTCT	265
OTGprimer7R	GGTGACTGATCCGATGAAGG	
OTGprimer8F	TCTCAGCTCGACAGTAGTTATTC	257
OTGprimer8R	ACTGCTCTCACTGTCCGAGt	

OTGprimer9F	CATTCAAGATGCAACACCTACAA	226
OTGprimer9R	CCGTAGTTGAGGGGAGGGgc	
OTGprimer10F	CACTACATACTGACCGCCAC	254
OTGprimer10R	TGTGATCTGAACTTTTCCAGCC	
OTGprimer11F	TGTAGGAATCAGAGCCCAATATT	263
OTGprimer11R	AGCAACCTCACTCTTGTCTG	
OTGprimer12F	GTTGAAGGATGATGGGGTGC	251
OTGprimer12R	GAGCTCTGTCTTCCTCTTCCT	
OTGprimer13F	CGATTGGTGGCTGACTCATG	82
OTGprimer13R	TGAGAATTGACAGCCCTCCA	
OTGprimer14F	AAAGGAGAGAAAGGAGATACTGG	178
OTGprimer14R	TTCATGGGCAAAGAATGTTCCC	
OTGprimer15F	TGTAACATGTTCATCTTCTCCCT	261
OTGprimer15R	GGTGACATAATTCCCATCCGAC	
OTGprimer16F	GATTTCCAGATGCATGGGTCC	277
OTGprimer16R	TGGCTTTTCTTCCAACAGGA	
OTGprimer17F	GTGGGTGACTGCTTGACTTT	138
OTGprimer17R	GACATGCCAAAAGGCCCTTG	

Table 3.2: Top candidate binding proteins from microarray

11DNR	13DNR	Both
ELAVL4	CDC42EP4	ANAPC11
TCEAL5	CSDA	ELAVL2
LMOD3	FNIP1	HNRPC
C13orf8	H1FX	INTS12
CAST	H2AFY2	LIN28
	HIST1H1A	NCL
	RPL29	RALY
	ZNF533	SAFB2
	PCBP2	SF3B4
		SFRS5
		SIAHBP1

Table 3.3: Gene ontology analysis using Panther bioinformatics.

	Down-regulated		Up-regulated	
Biological Processes	Fold Enrichment	*p*-value	Fold Enrichment	*p*-value
Macrophage activation	3.6	6.43×10^{-3}	--	--
Immune system process	2.04	2.58×10^{-7}	--	--
Response to stimulus	1.54	1.47×10^{-3}	--	--
mRNA splicing, via splicesosome	--	--	2.71	3.39×10^{-4}
mRNA processing	--	--	2.31	8.80×10^{-4}
Nucleobase-containing cmpd metabolic process	--	--	1.28	6.50×10^{-3}
Metabolic process	--	--	1.15	6.49×10^{-3}
Molecular Function	Fold Enrichment	*p*-value	Fold Enrichment	*p*-value
Extracellular matrix structural constituent	4.24	6.55×10^{-3}	--	--
Structural molecule activity	1.74	2.22×10^{-4}	--	--
Receptor activity	1.65	7.91×10^{-3}	--	--
RNA helicase activity	--	--	4.12	1.28×10^{-4}
Translation initiation factor activity	--	--	3.15	4.66×10^{-3}
mRNA binding	--	--	2.88	1.04×10^{-3}
Helicase activity	--	--	2.78	9.58×10^{-3}
Translation regulator activity	--	--	2.78	9.58×10^{-3}
Nucleic acid binding	--	--	2.72	8.76×10^{-3}
RNA binding	--	--	2.58	5.25×10^{-10}
Cellular Component	Fold Enrichment	*p*-value	Fold Enrichment	*p*-value
Extracellular matrix	--	--	2.56	5.25×10^{-3}
Extracellular region	--	--	2.43	1.56×10^{-5}

Figures: Chapter 3

Figure 3.1: The mTOR pathway

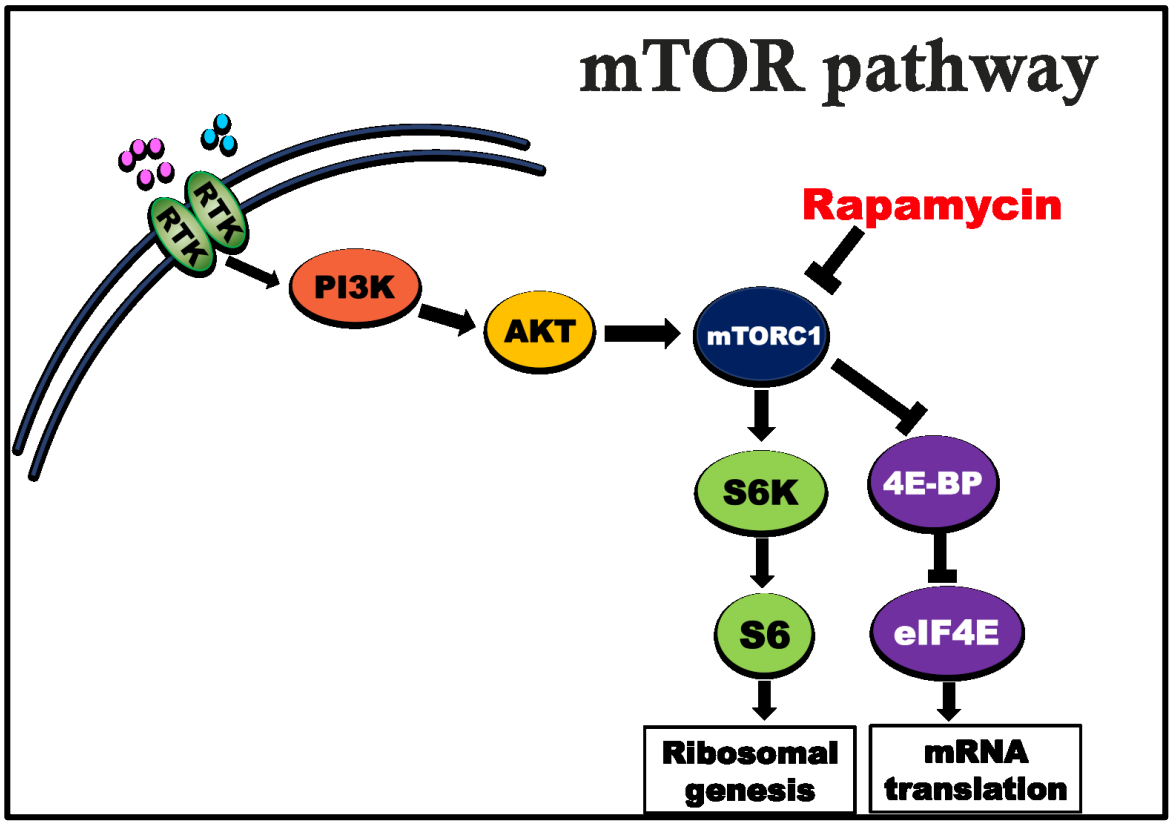

Figure 3.1: External stimuli such as nutrients and growth factors activate receptor tyrosine kinases (RTK), which then lead to the activation of downstream components, including phosphatidyl inositol-3-kinase (PI3K) complex and AKT. Once activated, mTORC1 phosphorylates S6K, which increases ribosome biogenesis, and 4EBP1, which increases 5' cap dependent mRNA translation. mTORC1 is highly sensitive to the immunosuppressant drug, Rapamycin.

Figure 3.2: Histogram distribution of Z scores

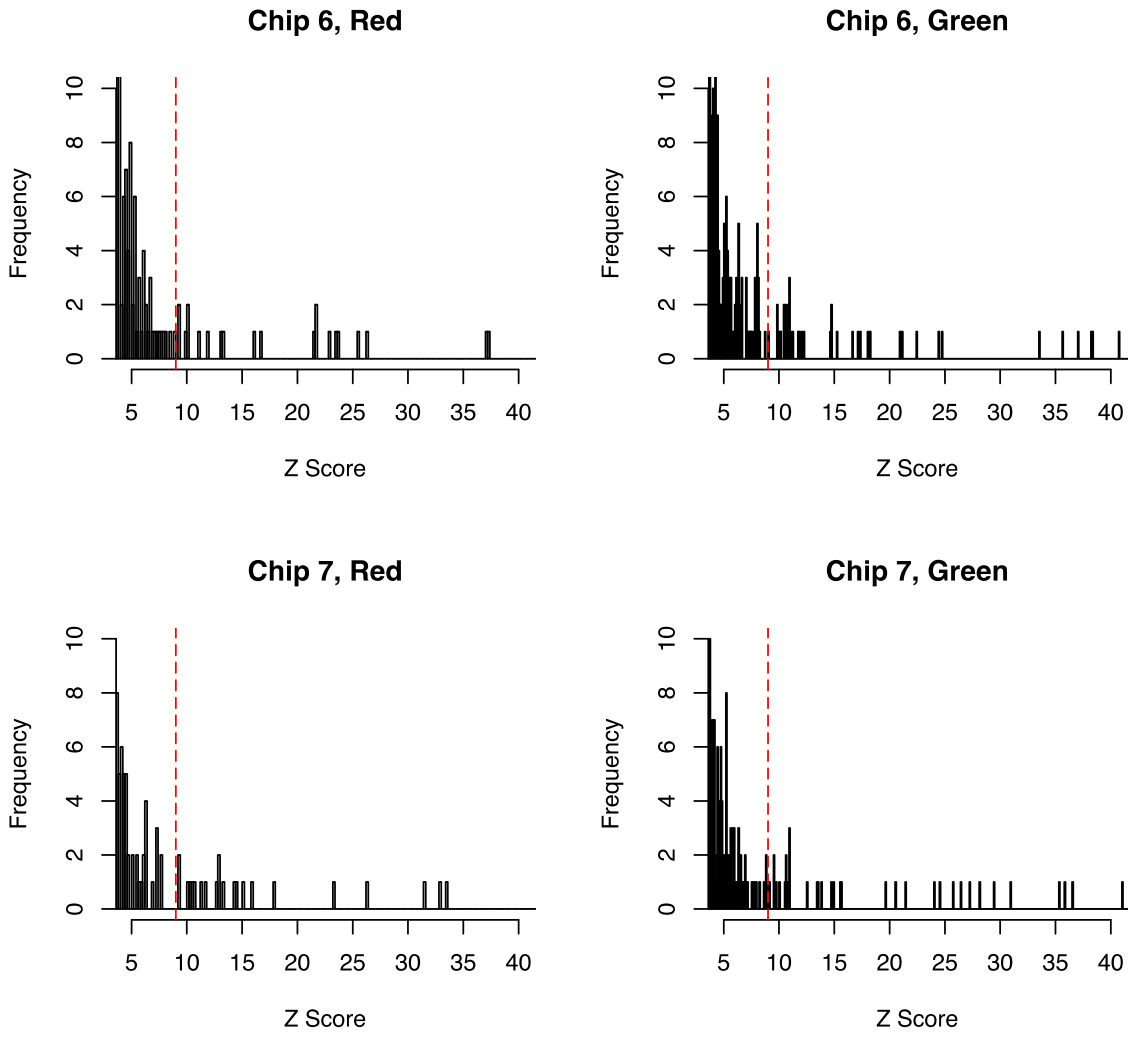

Figure 3.2: Histogram distribution of positive protein bindings to each of the alleles (Red fluorophore = 11DNR, Green fluorophore = 13DNR). Selected proteins with Z score > 9, as visualized by red dotted line.

Figure 3.3: Rapamycin series in cultured cells

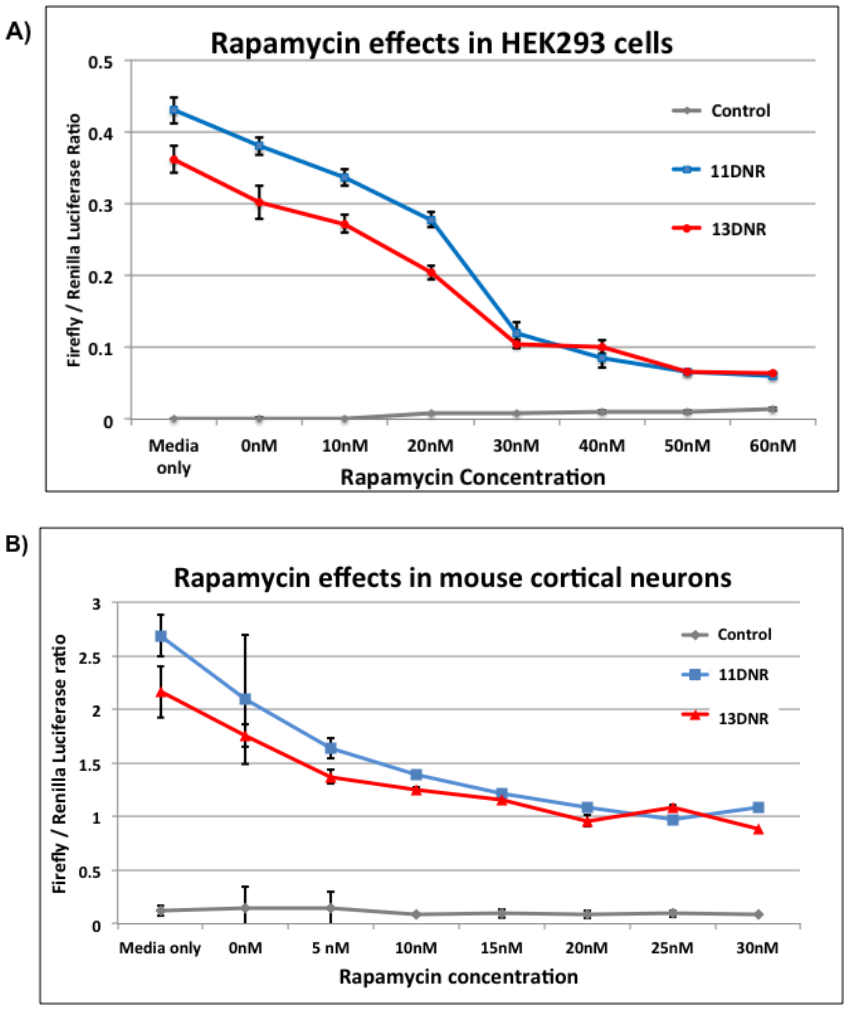

Figure 3.3: Exposure of Rapamycin to HEK293 cultured cells (A) and mouse primary cortical neurons (B) showed dose-dependent responses for both DNR alleles. In both cell types, increased Rapamycin led to decrease in luciferase activity, and complete inhibition at 30nM for HEK293 and 20nM for mouse primary cortical neurons.

Figure 3.4: EMSA of DNR allele interactions to proteins

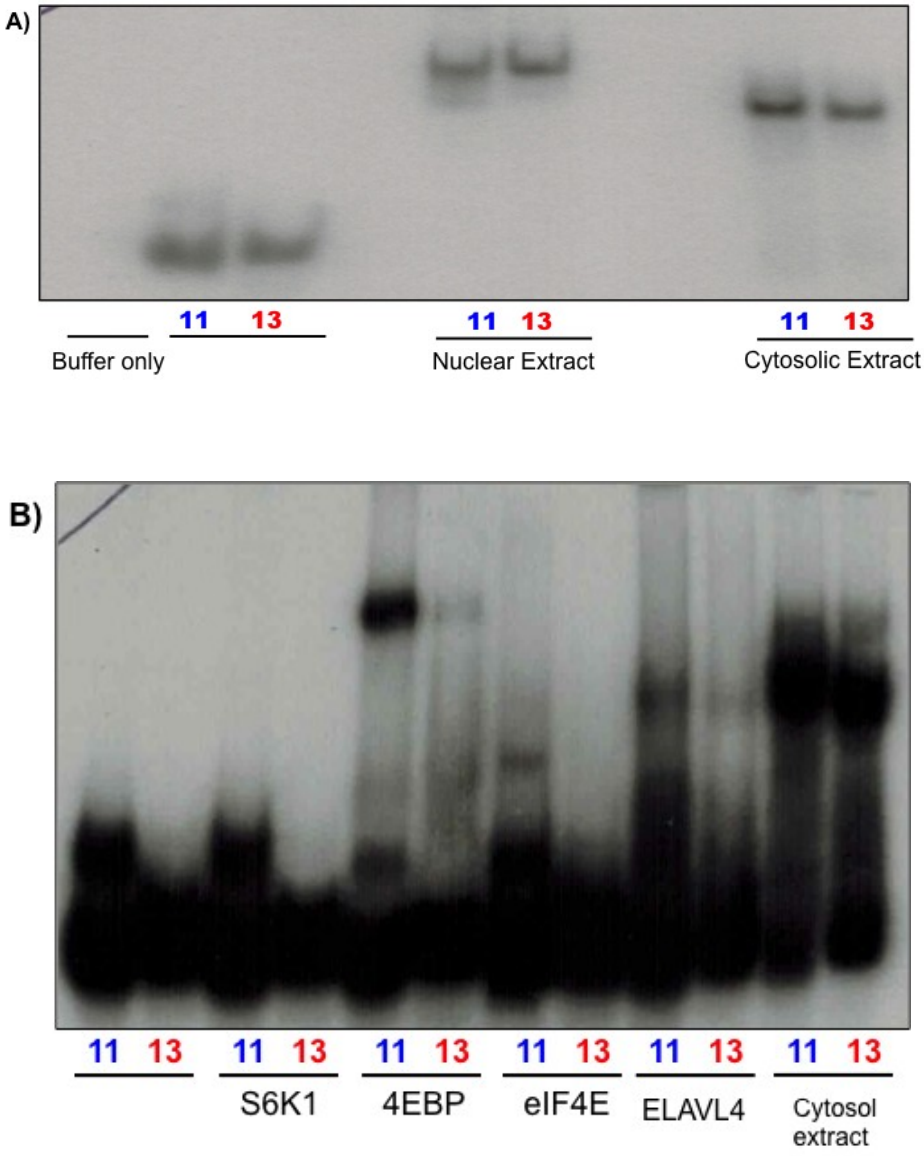

Figure 3.4: (A) Labeled DNR probes show clear shifts with the addition of the protein extracts, demonstrating the capacity to interact to whole proteins. **(B)** Labeled DNR probes and purified candidate binding proteins, confirming binding for 4EBP and eIF4E and ELAVL4. In all cases the binding was weaker, or eliminated for the 13DNR risk allele.

Figure 3.5: EMSAs resolving secondary structures in DNR probes

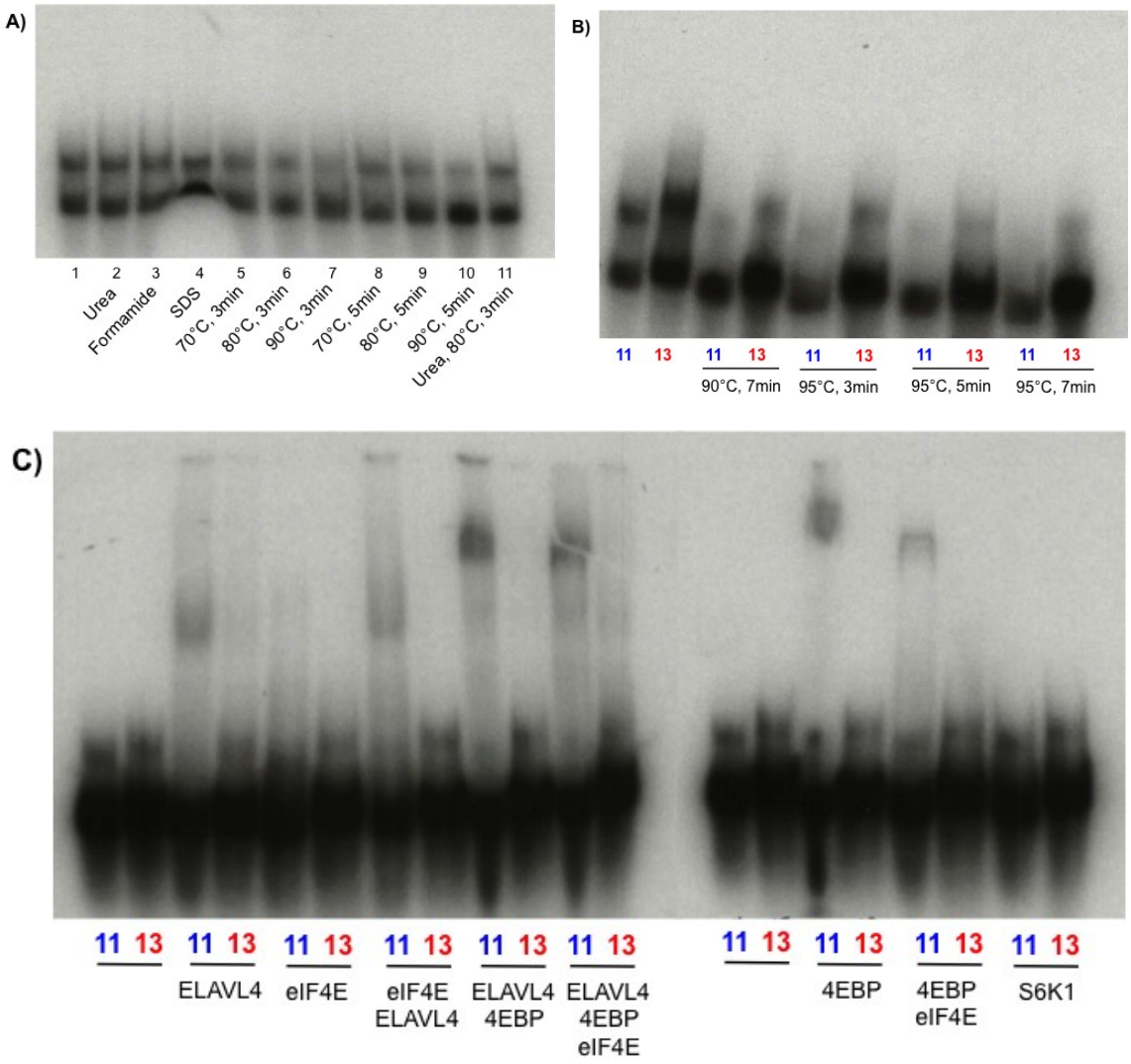

Figure 3.5: (A) Denaturing conditions at high temperatures for the 11DNR probe shows significant reduction of the secondary band. **(B)** New batch of DNR probes show secondary band for both alleles, which could also be diminished by heat treatment. **(C)** Validation of initial EMSA experiment and protein combinations with new DNR probes. No difference observed from our original results, despite additional band being present for both alleles.

Figure 3.6: Chromatographs of CRISPR-edited cell clones

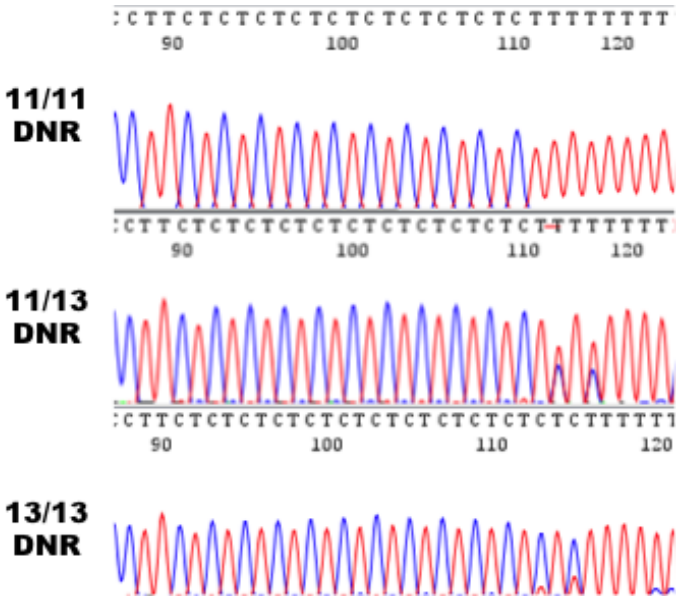

Figure 3.6: Chromatograph from Sanger sequencing confirming the genotypes of individual targeted clones at the DNR locus. Our targeting scheme yielded 4 clones that were 13DNR homozygous and 3 heterozygotes while the rest were not modified at the target site despite successful transfection, and therefore were appropriate controls.

Figure 3.7: Protein analysis of CRISPR-edited cell clones

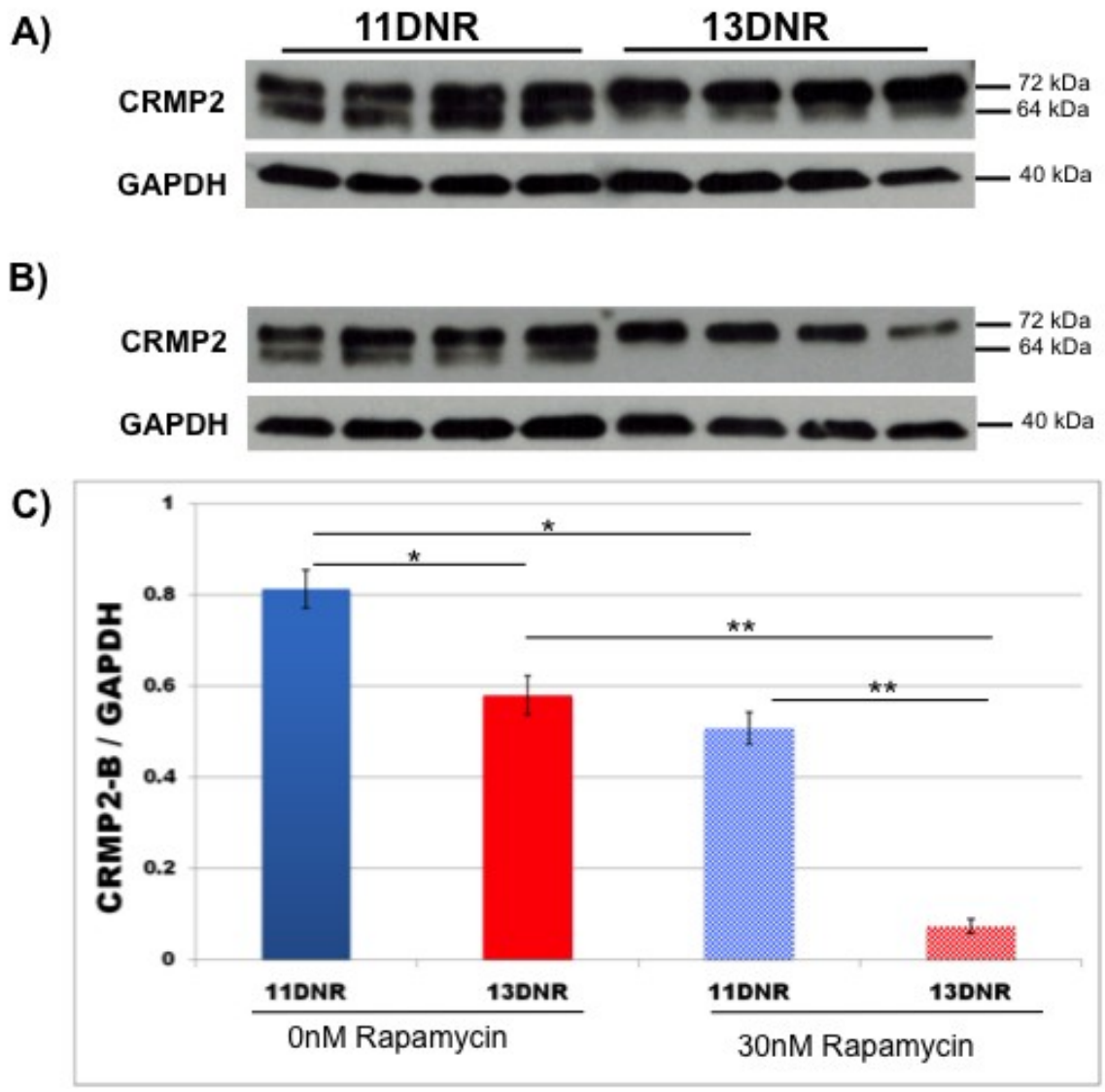

Figure 3.7: A) Western blots of targeted clones show significant differences in CRMP2B levels (64 kDa) between the two isogenic cell groups. **(B)** In the presence of 30 nM Rapamycin, CRMP2B is reduced for both cell groups, but most dramatically in the 13DNR cells. **(C)** Quantification of Western blots using Image J software. * $p < 0.001$; ** $p < 2 \times 10^{-5}$.

Figure 3.8: Morphological analysis of CRISPR-edited cell clones

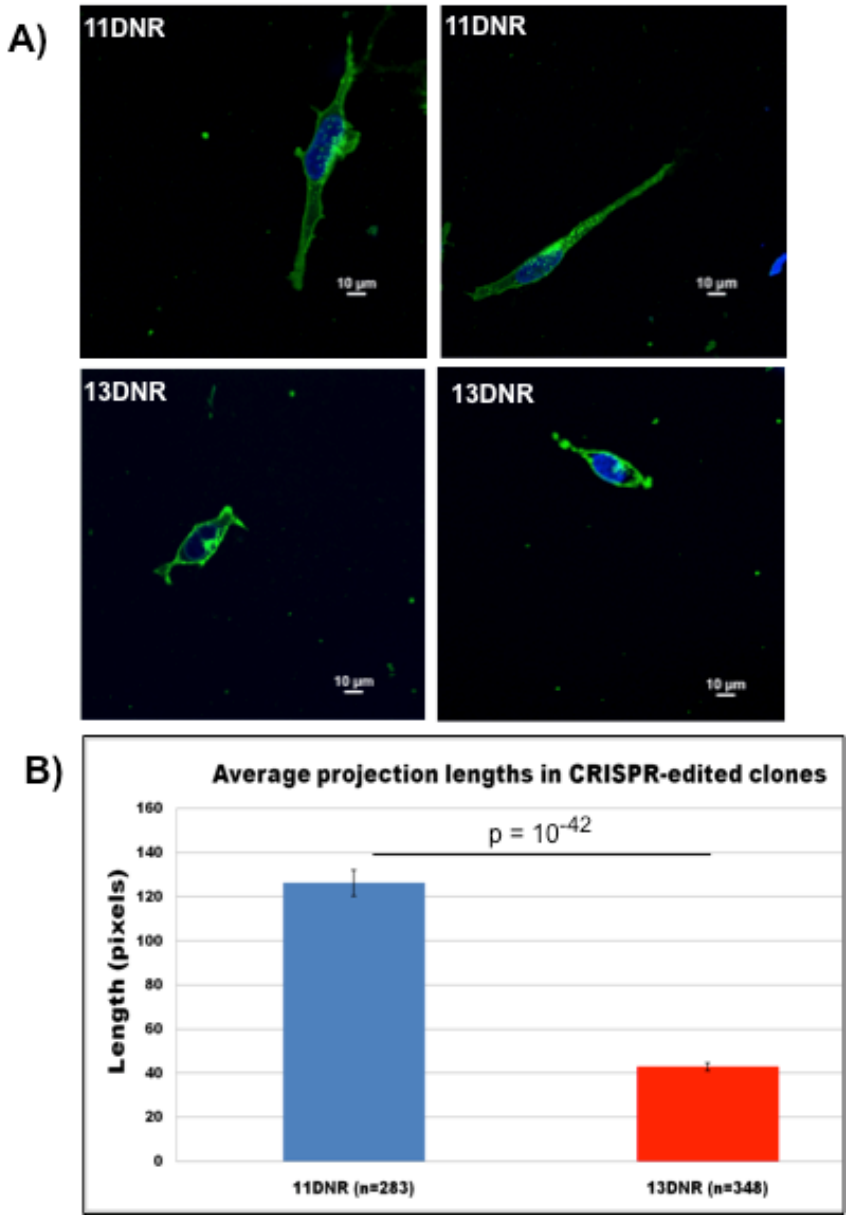

Figure 3.8: (A) Representative images of CRISPR-edited cells show morphological differences between the 11 and 13DNR cell groups. Cells stained for wheat germ agglutinin (green) and nuclei (blue). **(B)** Quantification of projection lengths between the two isogenic cell groups confirm a significant difference, with the 13DNR risk allele showing shorter projections.

Figure 3.9: RNA-seq QC analyses

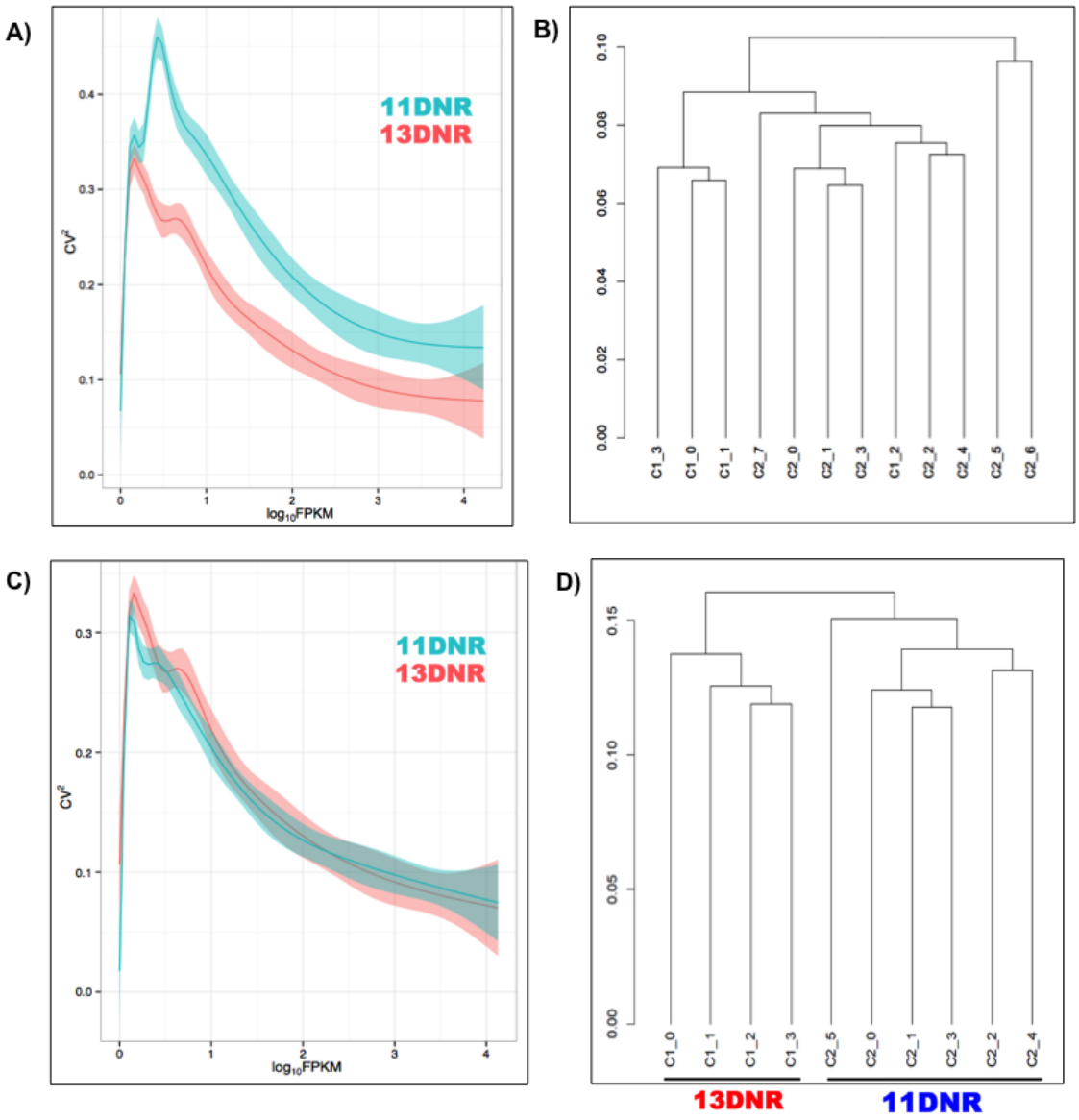

Figure 3.9: Variance profiles and dendrograms based on all genes. C1 is 13DNR, C2 is 11 DNR. **(A)** Variance profiles is significantly higher across all FPKM levels for the 11DNR. **(B)** Two 11DNR samples (C2_5, C2_6) cluster together and separate early from all other samples suggesting they are outliers. **(C)** Variance profiles for the two groups become similar after removal of the two outliers. **(D)** The dendogram now separates the

four 13DNR (C1) in one group, and six 11DNR (C2) in the other. Note that the numbering of the cell lines has now changed, C2_5 corresponds to previously C2_7.

Chapter 4: Creation of conditional *Dpysl2* knockout mice

Abstract

DPYSL2 encodes the CRMP2 protein that's important for axon specification and development. This gene is a good SZ candidate for several reasons: (i) it is located within the 8p21 SZ susceptibility locus with the strongest linkage signal coming from a SNP located a few kilobases at the 3' end of *DPYSL2*; (ii) its expression in the brain, and functions in axonal growth and neuronal migration are relevant to the study of SZ; (iii) it harbors the functional DNR that dramatically alters protein, morphology, and transcriptome levels *in vitro*; and, (iv), its regulation is linked to the mTOR pathway, which has been implicated in the pathology of neuropsychiatric diseases like SZ. To study the effects of this gene *in vivo*, we created a conditional knockout Dpysl2 mouse model that lacks the third exon of the gene in the forebrain. Preliminary behavioral tests in a small subset of heterozygous knockout mice (n = 10) show interesting increase in exploratory activity and reduction of anxiety levels. Additional behavioral testing and morphological analyses are necessary before we can confidently conclude the effects of *Dpysl2* loss of function on mouse phenotypes. Nevertheless, the conditional *Dpysl2* knockout model is a powerful tool that will serve as the foundation for future experiments that center on *Dpysl2* loss of function and SZ.

Introduction

In this chapter, I describe the creation of the *Dpysl2* knockout mouse model for the study of SZ. Modeling loss of function of a gene in an animal model has been an invaluable tool in the study of psychiatric diseases, as mice can model certain quantifiable facets of the disorders that mirror the human symptoms [91–94]. Some of these genetic associations can be observed to alter behaviors, physiological, and morphological traits [95–97]. As *DPYSL2* is highly conserved and has known functions in axonogenesis [16,19,98], we hypothesized that loss of function of *Dpysl2* in the mouse would result in phenotypic changes at the neuronal level and at the behavioral level.

To our knowledge at the time of this writing, there have been no publications on a knockout model of *Dpysl2* for the study of SZ. However, a targeted knock-in *DPYSL2* mouse model was reported via substitution of an alanine in the place of a serine residue (*Dpysl2*-S522A) that is normally phosphorylated by Cdk5 [19]. Homozygous knock-in mice, designated $Dpysl2^{ki/ki}$, were viable and their cortical cyto-architecture was normal. When these animals were bred to animals in which the *Dpysl1* gene was disrupted to produce $Dpysl1^{-/-}$; $Dpysl2^{ki/ki}$ double homozygote animals, the authors showed that cortical neurons had strikingly abnormal dendritic "curling" phenotypes. None of the single or double mutants have been characterized for any neurobehavioral

phenotypes. The lack of a knock-out *Dpysl2* mouse model as well as the lack of a comprehensive phenotypic characterization of these animal models prompted us to create a knock-out mouse model of *Dpysl2*.

To test the consequences of loss of function of *Dpysl2*, we took advantage of a commercially available knockout allele for *Dpysl2* (*Dpysl2*$^{tm2e(KOMP)Wtsi}$), which at the time of this study, was available in embryonic stem (ES) cells via the UC Davis Knockout Mouse Project (KOMP) Repository. According to KOMP, the targeting event in this construct creates targeted ES cell clones without conditional potential through a gene-trapping event that introduces early polyadenylation signals, which interrupt transcription of the *Dpysl2* gene (Fig. 4.1). As the targeting scheme involves sequences shared by all isoforms, removal of the floxed region is predicted to produce a truncated protein product of for *Dpysl2A*, *Dpysl2B,* and *Dpysl2C.* We planned to produce both heterozygous and homozygous targeted mice for functional testing.

Also known as B6 albino, albino B6, B6(Cg)-Tyrc-2J/J mice have the C57BL/6J background but carry a mutation in the tyrosinase gene, which, in the homozygous form, results in white coat coloring. Following implantation and gestation, resulting offspring that are chimeric are easily identified by the black/white coat color. This method facilitates efficiency in the creation of knockout mice on the same C57BL/6J

genetic background as the ES cells, bypassing the need for multi-generational backcrossing [99–101].

With the help of the Johns Hopkins Transgenic Core, we obtained a colony of mice with positive germline transmission. However upon breeding these *Dpysl2* $^{+/neo}$; *Tyr* $^{+/-}$ mice to B6.C-Tg(CMV-cre) to remove the neo cassette, we observed complete perinatal lethality. Following extensive Southern blotting and sequencing of the entire targeted cassette, we identified an additional loxP site that flanked the 3' end of exon 3. Though this was part of the original construct, KOMP reported this loxP was lost in the targeting event into ES cells, thus rendering the line non-conditional. Our sequencing studies showed that this additional loxP was retained and its presence confers conditional potential to the targeting allele. With this new knowledge, we revised our breeding scheme to remove the entire LacZ-neo cassette by mating to B6.C-Tg(ACTFLPe) mice. The progeny of this cross, *Dpysl2* $^{+/lox}$ were then mated to B6.C-Tg(Camk2a-cre) mice to produce mice with conditional knockout of *Dpysl2* in the pyramidal neurons. We generated both heterozygous and homozygous mutants of these conditional knockouts and have subjected a small cohort of these mice to baseline behavioral tests.

Materials and Methods

Targeting scheme

We obtained C57BL/6N ES cells for the Dpysl2 knockout allele (*DPYSL2*$^{tm2e(KOMP)Wtsi}$) from the KOMP repository (project ID CSD38021). The targeting construct has an SA-IRES-lacZ-pA cassette and a neo-pA cassette flanked by *LoxP* sites. Although the vector was designed to produce the knockout first allele, which has conditional potential, the only clones available were described as "tm2e," meaning the third 3' LoxP site was missing and contain the targeted non-conditional allele. The gene-trapping event in this allele introduces early polyadenylation signals, which interrupt transcription of the *Dpysl2* gene, and is therefore predicted to result in complete loss of function of both *Dpysl2A and Dpysl2B* (Fig. 4.1). The ES cell clones were confirmed by genotyping protocol to be heterozygous for the mutant allele.

Microinjection

Confirmation of these lines was done through long PCR across the 5' and 3' homology arm junctions and Southern blotting with probes specific for the 5' flank and for the 3' flank (Table 4.1). We next worked with the Hopkins Transgenic Core Laboratory for expansion of positive clones and karyotype analysis to rule out aneuploidy. The best growing ES clones passing these screens were then injected by the Johns

Hopkins Transgenic Core Laboratory, a service that yielded high-grade chimeric male mice on the C57BL/6 background heterozygous for the Tyrosinase deficiency for easy coat color screening.

Breeding scheme

We bred the black and white chimera mice that contained the knockout cells to albino C57BL/6 mice (JAX #000058) to obtain positive GLT progeny, which was indicated by a pure black coat color. Using PCR genotyping, we selected GLT positive mice heterozygous for the *Dpysl2* targeted allele as well as for the Tyrosinase mutation, *Dpysl2* $^{+/neo}$;*Tyr*$^{+/-}$. Through an additional cross to wild-type C57BL/6 (JAX #000664), we obtained a line of mice heterozygous for the Dpysl2 targeted allele with no Tyrosinase deficiencies, *Dpysl2* $^{+/neo}$;*Tyr*$^{+/+}$, henceforth abbreviated as *Dpysl2* $^{+/neo}$ (Fig. 4.2).

To remove the neo-resistance cassette, which has been reported to interfere with the targeting construct, we mated *Dpysl2* $^{+/neo}$ mice to transgenic CMV-cre mice (JAX # 006054), which ubiquitously express the enzyme cre recombinase under the transcriptional control of a human cytomegalovirus minimal promoter. Using PCR genotyping, we screened for progeny that showed successful deletion of the neo-pA cassette that was flanked by *LoxP* sites. Because the progeny of this cross showed perinatal lethality due to an undisclosed LoxP site at the 3' end of exon 3, we modified our breeding scheme to take advantage of the

conditional potential of this construct (Fig. 4.2). We mated *Dpysl2* $^{+/neo}$ mice to transgenic ACTFLPe mice (JAX #005703), which ubiquitously express the enzyme FLP1 recombinase under the transcriptional control of the human actin B promoter. Presence of the FLP1 recombinase deleted the SA-IRES-lacZ-pA cassette and the neo-pA cassette that are flanked by FRT sites to yield mice with only the third Dpysl2 exon flanked by loxP sites, denoted as *Dpysl2* $^{+/lox}$. We crossed *Dpysl2* $^{+/lox}$ to *Dpysl2* $^{+/lox}$ mice to obtain a homozygous *Dpysl2* $^{lox/lox}$ colony, with only loxP sites flanking the third exon on both alleles. As ubiquitous removal of the third exon likely contributed to perinatal deaths, we then mated *Dpysl2* $^{lox/lox}$ mice to transgenic Camk2a-cre mice (JAX #005359), which express tissue-specific Cre recombinase under the direction of the calcium/calmodulin-dependent protein kinase II alpha (Camk2a) promoter. With this breeding scheme, we derived mice with two sets of genotypes. The first were heterozygous *Dpysl2* $^{exon3+/-}$;CamCre.Tg mice that also expressed Cre recombinase in the forebrain, which knocked out the third exon only in the pyramidal cell layer in the hippocampus. The second set of mice from this cross were of the genotype *Dpysl2* $^{+/lox}$, which were heterozygous for the *Dpysl2* targeted allele but absent for the Cre recombinase, and therefore served as convenient wild-type littermate controls. (Fig. 4.3)

Genotyping assays

To confirm presence of the *Dpysl2* targeted allele, we used primers (DPYSL2_neoF & DPYSL2_ttr) recommended by KOMP that span the end of the neo cassette into the 3' homology arm (Table 4.1, Fig. 4.4). The 557 bp fragment indicated the presence of the targeting cassette. To confirm deletion of the neo cassette with CMV-Cre mice, we designed primers (Floxed_2F & Floxed_2R) that span from the end of the lacZ cassette to the 3' homology arm. The 2,419 bp fragment indicated the loxP sites along with the entire neo cassette were intact; however, absence of this fragment indicated the region was floxed (Table 4.1, Fig. 4.4). To confirm the integration of the targeting event into the mouse genome, we designed primers that tested the integration at the 5' and 3' end of the entire cassette. Integration at the 5' end used primers (5'_Dpy_int_F & 5'_Dpy_int_R) that resulted in a 5,240 bp fragment. Integration at the 3' end used primers (3'_Dpy_int_F & 3'_Dpy_int_R) that resulted in a 5,703 bp fragment (Table 4.1, Fig. 4.4).

To test for aberrant floxing due to the discovered third loxP site, we designed primers (Floxed_F1 & 3rdLoxP_R3) that spanned all three loxP sites. In this PCR, if the entire fragment was intact, the fragment size would be 8.6 kb. However, floxing combinations between the three loxP sites would result in smaller fragment sizes as follows: 6.6 kb if floxed between the first and second loxP sites; 7.6 kb if floxed between the

second and third loxP sites; 5.7kb if floxed between the first and third loxP sites (Table 4.1, Fig. 4.4).

Finally, to derive at the genotypes after crossing to ACTFLPe mice, we designed a double set of primers to reveal the genotype for both alleles. The first set of primers (lacZ_2F & lacZ_2R) span a 1.1 kb portion of the lacZ cassette, indicating presence of the targeted allele. The second set of primers (DPYSL2_F & DPYSL2_R) span from the 5' homology arm to the 3' homology arm, resulting in a 7.6 kb fragment for the targeted allele, and a fragment that's 564 bp for the wild-type allele. Because PCR preferentially amplifies smaller amplicons, the 7.6 kb fragment is not visualized in a gel. After crossing to the ACTFLPe mice, excision at the FRT sites will remove 6.9 kb of genomic DNA, leaving a smaller 673 bp fragment to amplify from the DPYSL2_F/R primers, and no amplification for the lacZ_2F/R primer set. Thus, this PCR reveals several genotypes: wild-type *Dpysl2*$^{+/+}$ (one 564 bp band), heterozygous *Dpysl2* $^{+/neo}$ (one 564 bp band, one 1.1 kb band), heterozygous *Dpysl2* $^{+/lox}$ (one 564 bp band, one 673 bp band), and homozygous *Dpysl2* $^{lox/lox}$ (one 673 bp band) (Table 4.1, Fig. 4.5).

To test the presence or absence of CMV-cre expression, we used primers Cre_1084 and Cre_1085. For Cam2ka-Cre, I used these primers CamCre_F and CamCre_R.

Southern blot

In order to investigate the cause of the perinatal deaths, we conducted southern blots to detect possible aberrant integration of the targeted cassette into the mouse genome. We designed three sets of probes specific to the 5' homology arm (Probe 1), the lacZ cassette (Probe 2), and the junction at the neo cassette and the 3' homology arm (Probe 3) (Fig. 4.6). For Probe 1, wild-type mice would show a 11.1 kb fragment, while heterozygous *Dpysl2* $^{+/neo}$ mice would show a 13.4 kb fragment in addition to the 11.1 kb fragment. Probe 2 is specific to the inserted cassette, and would only show one 13.4 fragment to indicate presence of the lacZ in *Dpysl2* $^{+/neo}$ mice. Similarly, Probe 3 shows as a 4.8 kb fragment only for the *Dpysl2* $^{+/neo}$ mice.

For southern blots, we used DNA from 18 mice: 3 *Dpysl2* $^{+/neo}$ parents; 4 *Dpysl2* $^{+/neo}$ pups with the targeted allele and no Cre; 10 *Dpysl2* $^{exon3\ +/-}$ pups with ubiquitous Cre expression; and, 1 *Dpysl2* $^{+/+}$ albino mouse as a negative control. The DNAs were digested with the restriction enzymes XhoI (NEB #R0146S) and AflII (NEB #R0520S) and loaded in a 0.9% agarose gel. The gel was denatured and transferred to a nylon membrane (Bio-Rad #1620177). Probes for the blot labeled with random primer labeling kits (Pharmacia Biotech #RPN1633), and P^{32} dCTP. The membrane was hybridized to the probe using salmon sperm DNA and hybridization solution at 42 C overnight. The membrane was

washed and exposed on film (Sigma #350370). After autoradiography, membranes were boiled in stripping solution and used for hybridization with other probes.

Behavioral tests

To gather preliminary behavioral data on our set of *Dpysl2* $^{exon3\ +/-}$ mice and the control littermates *Dpysl2* $^{+/lox}$, we opted for a suite of well-known baseline behavioral tests. This included the open field test for novelty-induced activity, Y-maze and spontaneous alternations for working memory, and the elevated plus maze (EPM) for anxiety levels. For these tests, we used mice 8-10 weeks old, 10 females and 3 males; however we excluded the 3 males from the end analysis. In addition, because testing anxiety can skew behavioral data, we performed the tests in order of increasing stress to the mouse, starting with open field, followed by spontaneous alternations, elevated plus maze, and ending with Y-maze. The genotypes of the mice were not revealed until after the data were collected. Brief descriptions of the tests are as follows:

Open field test: This test measures locomotor activity and anxiety induced by novel environments [102,103]. Each mouse was placed in a square chamber (16”x16”x16”) equipped with laser beams that recorded locomotor activity. After an acclimation period, we measured the frequency that a beam broke in 5-minute intervals. Mice were observed for a total of 45 minutes.

Spontaneous alternations: This test assesses working memory and tests the mouse's ability to alternate of to all three different arms consecutively [104,105]. For this, all three arms of the maze are open, and we manually recorded the number of triads (defined as entry into all three different arms in succession) in a period of five minutes. This was calculated as the number of triads divided by the maximum possible alternations (total alternations minus 2).

Y-maze: This test also evaluates hippocampal-dependent learning and working memory and is focused on the mouse's innate curiosity to explore new areas [106,107]. Mice were conditioned to explore the maze for 5 minutes with one random arm of the maze closed off. After 15 minutes, the mouse was returned to the maze with all three arms opened, and tested on the amount of time it spent in the previously closed arm. Entries and time spent in each arm was recorded via a camera.

Elevated Plus Maze: This test measures anxiety levels, for which high anxiety would manifest as more time spent in the closed arm [108–110]. The mouse was placed in the middle of the T-shaped contraption, and we recorded the time it spent in the open arm versus the closed arms of the maze for a total of 5 minutes.

Results

Ubiquitous knockout of Dpysl2 results in perinatal lethality

After obtaining heterozygous *Dpysl2* $^{+/neo}$ mice with positive germline transmission, we intended remove the neo cassette, as this has been reported to interfere with the genetic background. However, we observed a high instance of perinatal lethality in pups produced by crossing heterozygous *Dpysl2* $^{+/neo}$ mice to transgenic CMV-Cre mice. The perinatal deaths were not sex-specific, as deaths occurred for both males and female parental with the heterozygous *Dpysl2* $^{+/neo}$ genotype (Fig. 4.7). Necropsies and genotyping data were normal and did not point to a clear cause of death. We observed the deaths of the pups between postnatal day (P) 0 and P3, after which the entire litter was completely lost. Though the time of deaths varied, some pups that lived past P2 had the appearance of milk spots, suggesting that they were able to obtain nutrients from the mother. We observed bite wounds on some of the dead pups, but we could not ascertain whether the pups died first and were eaten, or the deaths were the result of infanticide by the mother. However, the fact that we observed the same pattern of mortality in many successive matings with the same parents indicated the deaths were not due to infanticide. Furthermore, we observed no evidence of perinatal deaths in the litters resulting from the matings between the heterozygous *Dpysl2* $^{+/neo}$ mice to wild-type mice, and those from matings between the

CMV-Cre to wild-type mice. This line of evidence suggested the deaths were specific to the *Dpysl2* targeted allele.

Southern blots using probes specific to the *Dpysl2* targeting cassette revealed abnormal DNA sequence pattern in the targeting construct for the dead pups (Fig. 4.8A, B, C). Sequence results for the ES cells and one chimeric mouse revealed the presence of a third LoxP site, oriented in the same direction as the other known LoxP recombination sites, located at the 3' end of the homology arm. In three samples from mice that were heterozygous for the targeting allele and expressing the Cre enzyme, the Sanger traces showed deletion of large DNA segment between the first and third loxP sites. This was confirmed with the southern blot, which showed a band close to 5.7 kb and concordant with this pattern of floxing (Fig. 4.8B). From these results, we suspected deletion of the third exon of *Dpysl2*, even in heterozygous state, was incompatible with life. This hypothesis was confirmed, as once we amended the breeding scheme to create conditional *Dpysl2* knockouts with Cam2ka-Cre mice, the deaths were no longer observed. Genotyping PCRs for mice obtained from this modified scheme showed the expected patterns of genomic deletions: *Dpysl2*$^{+/+}$ (one 564 bp band), heterozygous *Dpysl2* $^{+/neo}$ (one 564 bp band, one 1.1 kb band), heterozygous *Dpysl2* $^{+/lox}$ (one 564 bp band, one 673 bp band), and homozygous *Dpysl2* $^{lox/lox}$ (one 673 bp band) (Fig. 4.5).

Baseline behavioral tests for Dpysl2 exon3 +/-

As a preliminary investigation into the behavioral phenotypes of *Dpysl2* conditional knockouts, we subjected a small cohort of female heterozygous *Dpysl2* exon3 +/- mice (n = 3) and the wild-type littermates (n = 7) to four baseline behavioral tests. The first test, the open field, measured the mouse's level of activity at the peripheral edges of the chamber and the center of the chamber as a proxy for anxiety levels induced by novel environment. Furthermore, this paradigm can also inform about the animals' innate curiosity to explore novel environments. For this this test, central activity, peripheral activity, and the combination of both central and peripheral activity were non-significant ($p > 0.8$). However, rearing, which is a measure of exploratory activity, was significant ($p=0.016$) with greater rearing activity for heterozygous knockouts. In subsequent tests, the mice showed no statistically significant differences for the spontaneous alternation ($p = 0.79$) or the Y-maze ($p = 0.37$) (Fig. 4.9 A, B). The EPM test showed mild significance ($p = 0.018$) with heterozygous knockouts displaying less anxiety in the open arm (Fig. 4.9 C). We conclude from this that the sample sizes are too small for any steadfast interpretations and that additional testing in larger sample sizes are necessary.

Discussion

In this part of my thesis project, I aimed to explore the functional consequences of *Dpysl2* loss of function in a mouse model. Although SZ is a human disease, mouse models can provide insight to certain facets of SZ-related symptoms, the results of which may point to genetic or other externally-induced alterations in the brain. Because *Dpysl2* has vital functions in a biologically relevant organ, the brain, we suspected knocking out this gene in the mouse would alter the animals' phenotype at some level. Indeed, our first breeding scheme to remove the neo cassette with CMV-Cre mice showed that *Dpysl2* is essential for life, as knocking out just one copy of this gene resulted in perinatal lethality. That the deaths occurred between P0-P3 suggested early developmental failure, though the exact mechanism is yet unknown. Pups that died were observed to have normal parturition, normal feeding, and no other obvious anatomical defects. CRMP2 expression has been shown to be the highest during the first postnatal week when maturation of neurons and synaptic connections are most active [15,111]. Thus, it is possible that the deaths could point to a mechanism of aberrant neurogenesis brought on by deletion of *Dpysl2*. However, further necropsies and molecular analyses are needed to rule out other causes of death and pinpoint the exact role of *Dpysl2* in the death phenotype.

After discovering the presence of the third loxP site and the newly found conditional potential of this construct, we generated a line of heterozygous *Dpysl2* $^{exon3\ +/-}$ mice that had deletion of the gene in the pyramidal cells of the hippocampus. In the preliminary behavioral tests, these mice showed varying significance in phenotype alterations as compared to their wild-type littermates. In the paradigms that test spatial working memory and hippocampal-dependent learning, there was no statistical difference between the heterozygous mutants and the controls. However, the heterozygous conditional knockouts displayed statistically significant differences in the open-field test specific to rearing activity, and in the elevated plus maze test. It's interesting that the significance overlap for these paradigms, as both measure the animal's anxiety induced by open space and novel environments. That heterozygous mutants displayed less anxiety and increased exploratory drives seem counterintuitive but could suggest possible novel connections between the gene and brain behaviors. Nonetheless, these results are preliminary at best, as my sample size was extremely limited. Additional testing in more mice are needed before we can confidently conclude the effects of *Dpysl2* loss of function in mouse behaviors.

In summary, we've developed a conditional *Dpysl2* knockout mouse model that has the potential to be a powerful tool in the study of SZ, and also genetics in general. To our knowledge at the time of this writing, no such mouse model exists for this gene in this background

strain. This model serves as the foundation for other future experiments that center on *Dpysl2* loss of function. In extended behavioral phenotyping of these heterozygous and homozygous knockout mice, we can conclude if and how loss of *Dpysl2* alters behaviors, and whether these changes could be proxy phenotypes to SZ-related symptoms. Morphological analysis of the brain at the gross and microscopic level could help us understand how *Dpysl2* influences axonogenesis and brain development. It would also be informative to harvest the brains of heterozygous and homozygous mice at various stages of development to examine the consequences of *Dpysl2* LOF on gene expression profiling, CNS development, and on neuronal morphology. In particular, this mutant mouse line offers the opportunity to culture primary cortical neurons carrying the loss-of-function mutation to study the role of *Dpysl2* in neural polarity and axonogenesis in vitro. This set up could also allow the application of Rapamycin to assess *Dpysl2* expression and mTOR inhibition in a more relevant cell model. Though the genetic design is different, it would be interesting to explore whether mTOR inhibition will accentuate these cells' phenotypic manifestations similar to what we observed with HEK293 isogenic cells. Finally, if the *Dpysl2* knockout mice do show phenotypic changes consistent with our expectations of the gene, the model provides an excellent opportunity to test various ways to rescue these alterations *in vivo*, including a host of small molecule compounds.

Tables: Chapter 4

Table 4.1: List of primers

PCR name	Primer ID	Sequence	Product (bp)
CSD-Lac	CSD-lacF	GCTACCATTACCAGTTGGTCTGGTGTC	690
	CSD-Dpysl2-ttR	AACTTACACGAGGCAGCAAATTCCC	
Neo	CSD-neoF	GGGATCTCATGCTGGAGTTCTTCG	557
	CSD-Dpysl2-ttR	AACTTACACGAGGCAGCAAATTCCC	
5' integration	5'_Dpy_integrxn_F	AGATTCCAGGCAGGGAGAAT	5240
	5'_Dpy_integrxn_R	GCCGCTTGTCCTCTTTGTTA	
3' integration	3'_Dpy_integrxn_F	GGATTCAGGGCAGTCTGAAA	5703
	3'_Dpy_integrxn_R	TGGACAAAACCTTCCAATCC	
Nested LacZ2	DPY_lacZ2_F	ACTCGCTCACATTTAATGTTGATGAAA	1149
	DPY_lacZ2_R	GGAAGGGCTGGTCTTCATCCACGCGC	
Floxed_2	Floxed_F_2	CTGAATATCGACGGTTTCCATATG	2419
	Floxed_R_2	CCGCCTACTGCGACTATAGAGATA	
Floxing at 3rd loxP	Floxed_F_1	GAATTCCAGCTGAGCGCCGGTC	Variable
	3rdLoxP_R3	AACCTCAGGTATAATACAGAGGG	
Nested Dpy	CSD-Dpysl2-F	CCACATCTTCGGAGGTATAGGCAGG	Variable
	CSD-Dpysl2-ttR	AACTTACACGAGGCAGCAAATTCCC	
CMV-Cre	112F	CTAGGCCACAGAATTGAAAGATCT	112
	112R	GTAGGTGGAAATTCTAGCATCATCC	
	Cre_1084	GCGGTCTGGCAGTAAAAACTATC	329
	Cre_1085	GTGAAACAGCATTGCTGTCACTT	

CamK2a-Cre	CamCre_F	TGCCACGACCAAGTGACAGCAATG	377
	CamCre_R	ACCAGAGACGGAAATCCATCGCTC	
Tyrosinase	Tyr_F	ATAGTCACTCCAGGGGTTGC	584
	Tyr_R	CAATGGGTCAACACCCAT	

Figures: Chapter 4

Figure 4.1: Dpysl2 targeting vector

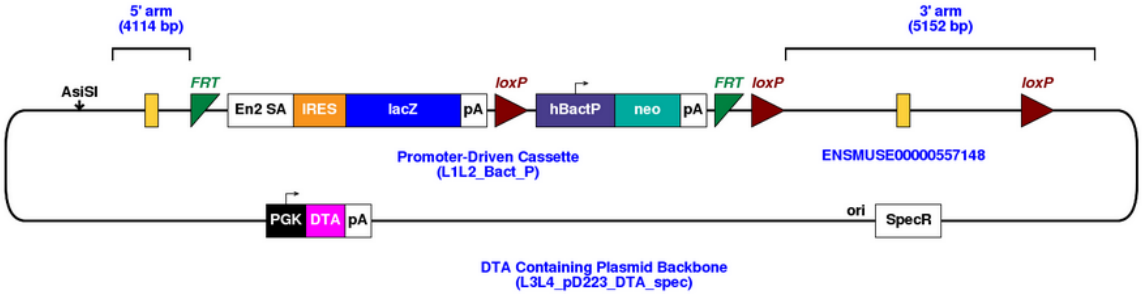

Figure 4.1: As described by KOMP, targeting vector originally contained 3 loxP sites. However, it was reported that the third 3' loxP site was lost in the recombination event, and so the construct works by early polyadenylation signals that terminate the *Dpysl2* transcript.

Figure 4.2: Initial breeding scheme to CMV-Cre mice

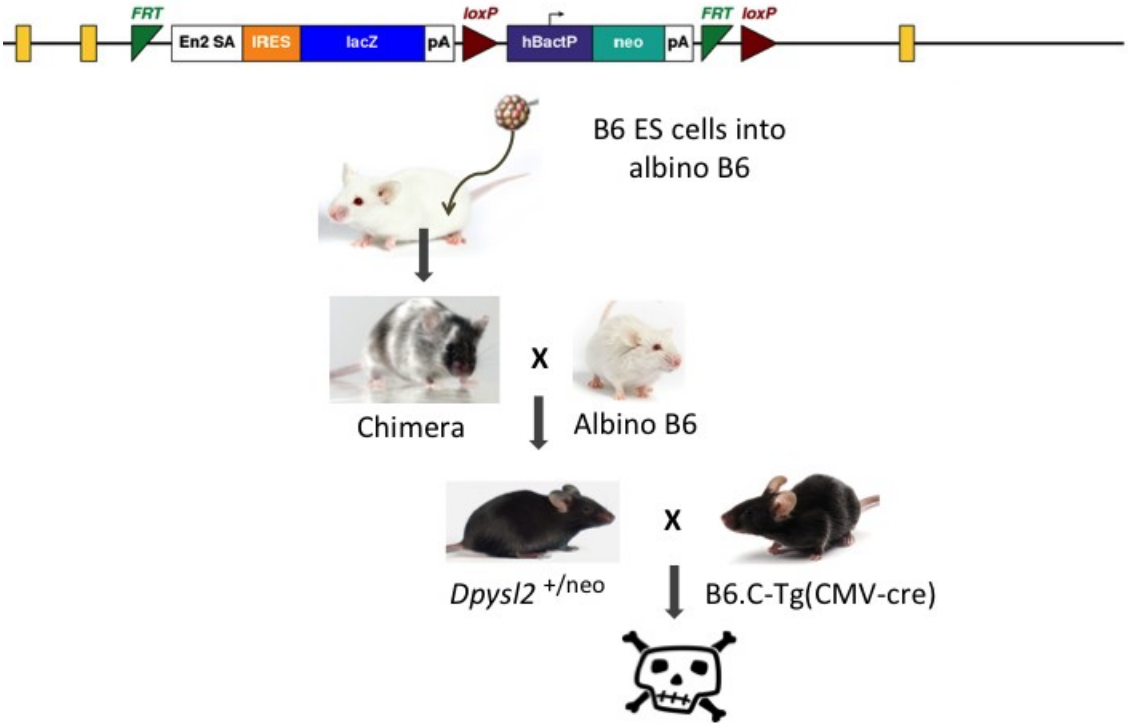

Figure 4.2: C57BL/6 derived ES cell clones were expanded, karyotyped, and microinjected into the blastocyst of a pseudopregnant albino black6 female. The high-grade chimeric male mice were then mated to albino Black 6 mice to obtain mice with positive germline transmission (Dpysl2 $^{+/neo}$). When these heterozygous mice were mated to CMV-cre mice, the resulting pups died perinatally.

Figure 4.3: Revised breeding scheme to produce conditional Dpysl2 knockouts

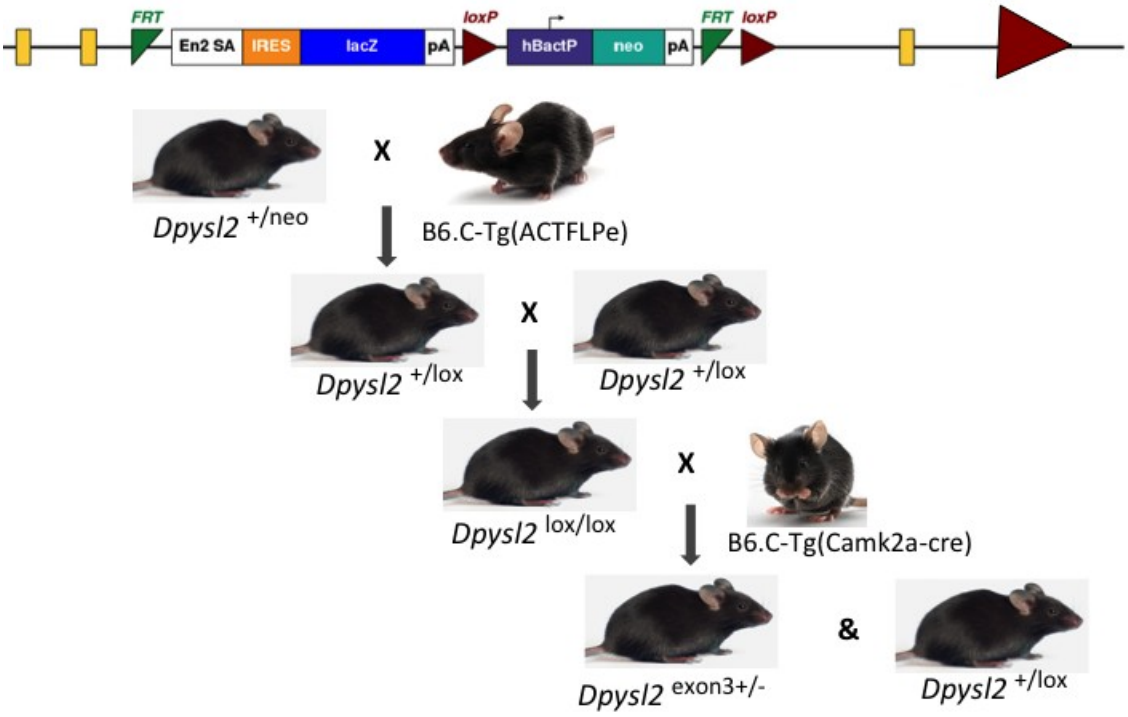

Figure 4.3: After learning about the intact third loxP site, we mated *Dpysl2* $^{+/neo}$ mice to FLP-expressing mice to remove the entire lacZ and neo cassette, leaving only the third exon flanked by loxP sites. Resulting pups were crossed to each other to obtain homozygous floxed mice, which were then mated to Camk2a-Cre mice to obtain conditional heterozygous knockouts (*Dpysl2* $^{exon3+/-}$) and wild-type littermate controls (*Dpysl2* $^{+/lox}$).

Figure 4.4: PCR scheme to identify aberrant floxing

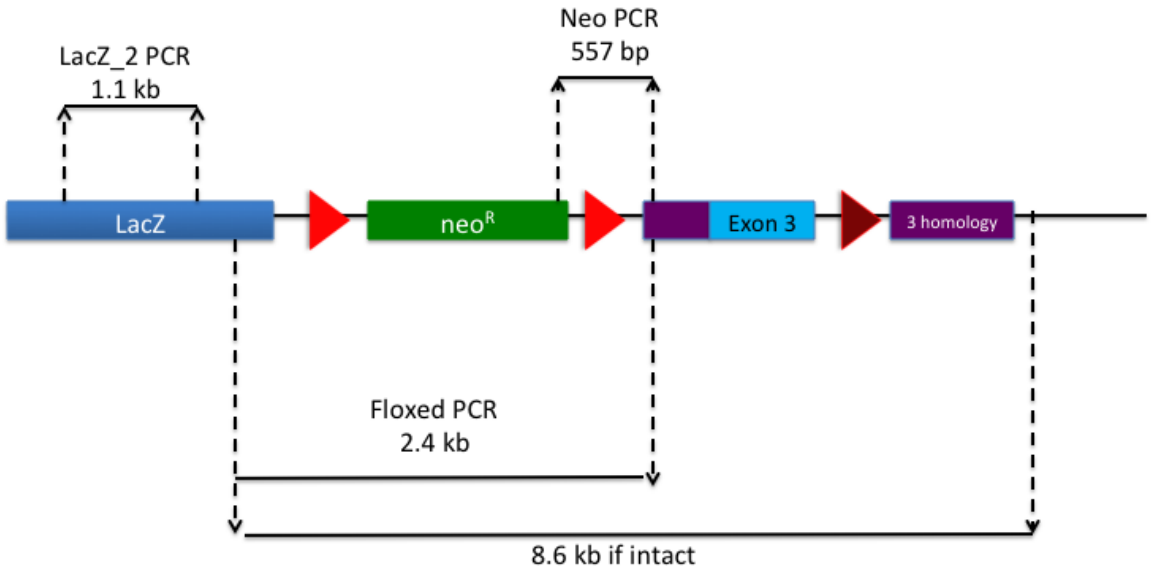

Figure 4.4: To identify the cause of aberrant floxing, I designed primers that tested the pattern of genomic deletions in mice. LoxP sites indicated by red triangles. The lacZ and neo PCR primers demonstrated insertion of the cassette in the heterozygous *Dpysl2* $^{+/neo}$ mice. Floxed PCR primers amplified 2.4 kb without deletion by Cre. In the presence of Cre, 6.6 kb would result if deletion occurred at the first and second LoxP; 7.6 if deletion occurred at the second and third LoxP; and ~5.7 kb if deletion occurred at the first and third LoxP.

Figure 4.5: PCR genotyping scheme

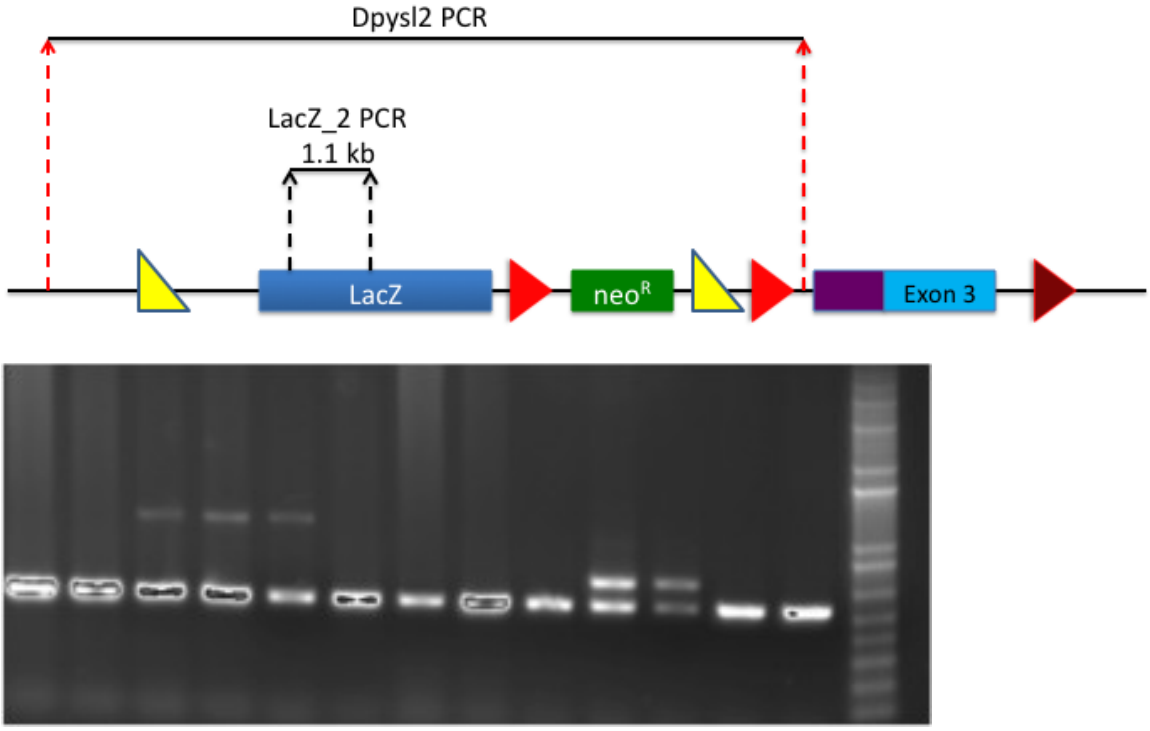

Figure 4.5: PCR genotyping scheme from matings to Camk2a-Cre mce to identify heterozygous targeted mice: *Dpysl2* $^{+/neo}$ (one 564 bp band, one 1.1 kb band); heterozygous conditional knockouts: *Dpysl2* $^{+/lox}$ (one 564 bp band, one 673 bp band); homozygous conditional knockouts: *Dpysl2* $^{lox/lox}$ (one 673 bp band), and wild-type controls: *Dpysl2*$^{+/+}$ (one 564 bp band). FRT sites are denoted as yellow triangles, and LoxP sites are denoted as red triangles.

Figure 4.6: Southern blot scheme

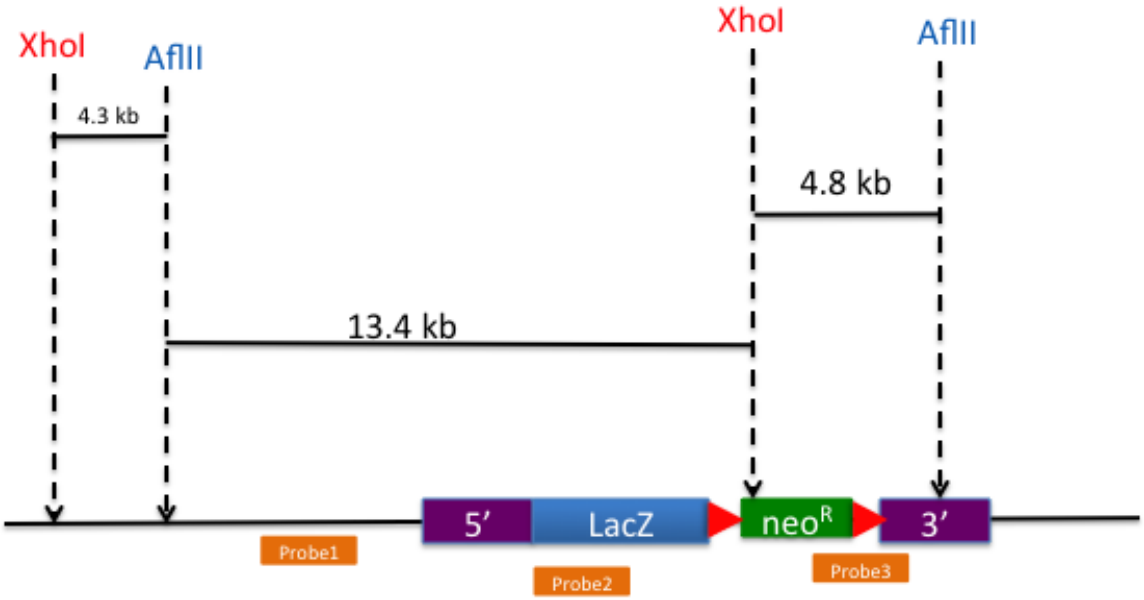

Figure 4.6: For Probe 1, which was specific to the 5' homology arm, wild-type mice would show a 11.1 kb fragment, while heterozygous *Dpysl2* $^{+/neo}$ mice would show a 13.4 kb fragment in addition to the 11.1 kb fragment. Probe 2 was specific to the inserted LacZ cassette, and would only show one 13.4 fragment to indicate presence of the lacZ in *Dpysl2* $^{+/neo}$ mice. Similarly, Probe 3, specific to the junction at the neo cassette and the 3' homology arm, would reveal a 4.8 kb fragment only for the *Dpysl2* $^{+/neo}$ mice.

Figure 4.7: Offspring deaths from matings to CMV-Cre mice

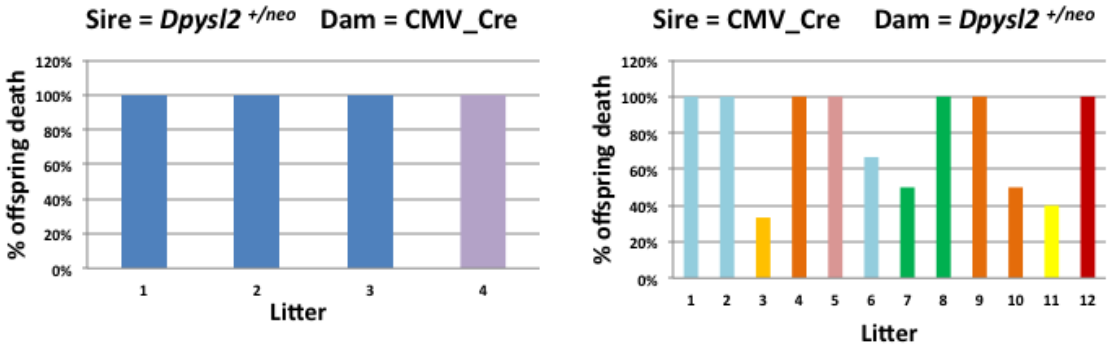

Figure 4.7: Perinatal deaths occurred when *Dpysl2* $^{+/neo}$ mice were crossed to CMV-Cre mice, regardless of the sex of the parents. Different colors indicate different dams.

Figure 4.8: Southern blots

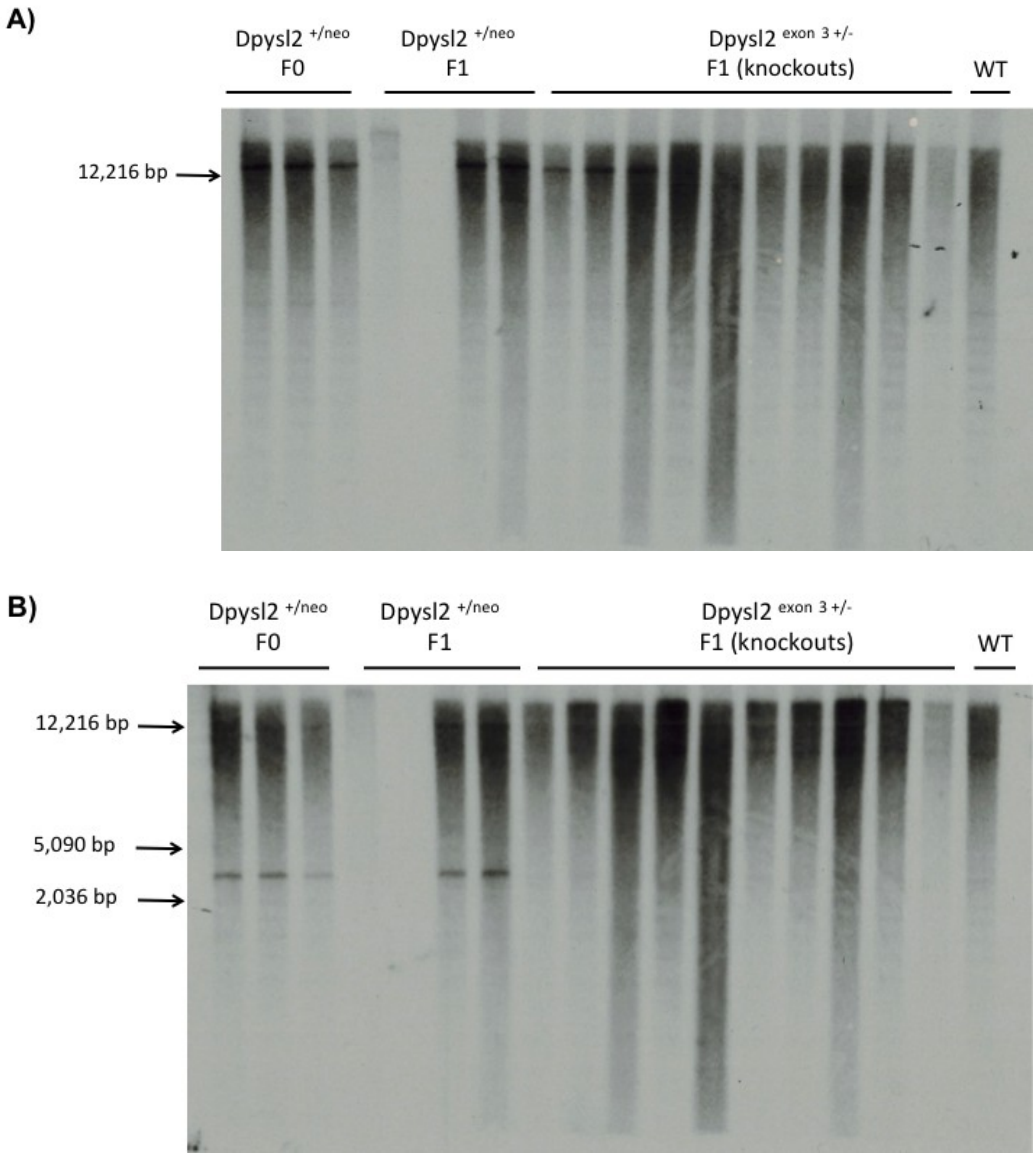

Figure 4.8: (A) Southern blots for the LacZ probe 1 confirmed heterozygous F0, F1 *Dpysl2* $^{+/neo}$ mice without Cre showed a 13.4 kb fragment, while the wild-type (WT) mice showed a 11.1 kb fragment. Some F1 heterozygous mice with Cre expression showed differences in band patterns compared to the parentals. **(B)** Blots for neo probe 2,

specific to the inserted cassette, showed a 4.8 kb fragment only for the *Dpysl2* $^{+/neo}$ mice, while those with Cre did not have this band.

Figure 4.9: Dpysl2 knockout behavioral tests

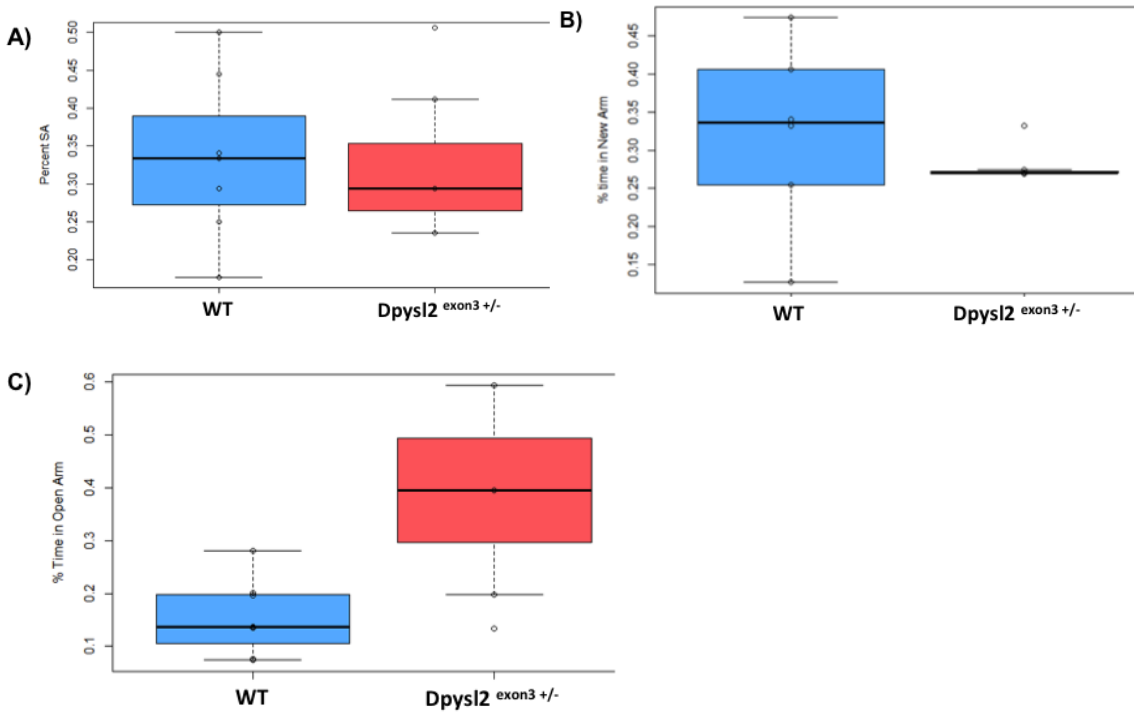

Figure 4.9: Baseline behavioral tests in conditional heterozygous knockout mice (n=3) compared to wild-type mice (n=7). **(A)** Spontaneous alternation paradigm showed no significant differences for memory deficits in either genotypes. **(B)** Y-maze paradigm did not show significant learning deficits in the heterozygous mutants. **(C)** Elevated plus maze showed heterozygous mutants spent more time in the open arm versus the closed arm, a measure of decreased anxiety ($p = 0.018$).

Bibliography

1 Eaton WW. Epidemiology of schizophrenia. *Epidemiol Rev* 1985; **7**: 105–126.

2 Wu EQ, Birnbaum HG, Shi L, Ball DE, Kessler RC, Moulis M *et al.* The economic burden of schizophrenia in the United States in 2002. *J Clin Psychiatry* 2005; **66**: 1122–1129.

3 Tsuang MT, Gilbertson MW, Faraone SV. The genetics of schizophrenia. Current knowledge and future directions. *Schizophr Res* 1991; **4**: 157–171.

4 Cardno AG, Gottesman II. Twin studies of schizophrenia: from bow-and-arrow concordances to star wars Mx and functional genomics. *Am J Med Genet* 2000; **97**: 12–17.

5 Gottesman II, Bertelsen A. Confirming unexpressed genotypes for schizophrenia. Risks in the offspring of Fischer's Danish identical and fraternal discordant twins. *Arch Gen Psychiatry* 1989; **46**: 867–872.

6 Kringlen E, Cramer G. Offspring of monozygotic twins discordant for schizophrenia. *Arch Gen Psychiatry* 1989; **46**: 873–877.

7 Badner JA, Gershon ES. Meta-analysis of whole-genome linkage scans of bipolar disorder and schizophrenia. *Mol Psychiatry* 2002; **7**: 405–411.

8 Lewis CM, Levinson DF, Wise LH, DeLisi LE, Straub RE, Hovatta I *et al.* Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: Schizophrenia. *Am J Hum Genet* 2003; **73**: 34–48.

9 Pulver AE, Lasseter VK, Kasch L, Wolyniec P, Nestadt G, Blouin JL *et al.* Schizophrenia: a genome scan targets chromosomes 3p and 8p as potential sites of susceptibility genes. *Am J Med Genet* 1995; **60**: 252–260.

10 Blouin JL, Dombroski BA, Nath SK, Lasseter VK, Wolyniec PS, Nestadt G *et al.* Schizophrenia susceptibility loci on chromosomes 13q32 and 8p21. *Nat Genet* 1998; **20**: 70–73.

11 Pulver AE, Mulle J, Nestadt G, Swartz KL, Blouin JL, Dombroski B *et al.* Genetic heterogeneity in schizophrenia: stratification of genome scan data using co-segregating related phenotypes. *Mol Psychiatry* 2000; **5**: 650–653.

12 Holmans PA, Riley B, Pulver AE, Owen MJ, Wildenauer DB, Gejman PV *et al.* Genomewide linkage scan of schizophrenia in a large multicenter pedigree sample using single nucleotide polymorphisms. *Mol Psychiatry* 2009; **14**: 786–795.

13 Fallin MD, Lasseter VK, Liu Y, Avramopoulos D, McGrath J, Wolyniec PS *et al.* Linkage and association on 8p21.2-p21.1 in schizophrenia. *Am J Med Genet Part B Neuropsychiatr Genet Off Publ Int Soc Psychiatr Genet* 2011; **156**: 188–197.

14 Arimura N, Menager C, Fukata Y, Kaibuchi K. Role of CRMP-2 in neuronal polarity. *J Neurobiol* 2004; **58**: 34–47.

15 Charrier E, Reibel S, Rogemond V, Aguera M, Thomasset N, Honnorat J. Collapsin response mediator proteins (CRMPs): involvement in nervous system development and adult neurodegenerative disorders. *Mol Neurobiol* 2003; **28**: 51–64.

16 Crews L, Ruf R, Patrick C, Dumaop W, Trejo-Morales M, Achim CL *et al.* Phosphorylation of collapsin response mediator protein-2 disrupts neuronal maturation in a model of adult neurogenesis: Implications for neurodegenerative disorders. *Mol Neurodegener* 2011; **6**: 67.

17 Ip JPK, Fu AKY, Ip NY. CRMP2: functional roles in neural development and therapeutic potential in neurological diseases. *Neurosci Rev J Bringing Neurobiol Neurol Psychiatry* 2014; **20**: 589–598.

18 Quach TT, Honnorat J, Kolattukudy PE, Khanna R, Duchemin AM. CRMPs: critical molecules for neurite morphogenesis and neuropsychiatric diseases. *Mol Psychiatry* 2015; **20**: 1037–1045.

19 Yamashita N, Ohshima T, Nakamura F, Kolattukudy P, Honnorat J, Mikoshiba K *et al.* Phosphorylation of CRMP2 (collapsin response mediator protein 2) is involved in proper dendritic field organization. *J Neurosci Off J Soc Neurosci* 2012; **32**: 1360–1365.

20 Edgar PF, Douglas JE, Cooper GJ, Dean B, Kydd R, Faull RL. Comparative proteome analysis of the hippocampus implicates chromosome 6q in schizophrenia. *Mol Psychiatry* 2000; **5**: 85–90.

21 Johnston-Wilson NL, Sims CD, Hofmann JP, Anderson L, Shore AD, Torrey EF *et al.* Disease-specific alterations in frontal cortex brain proteins in schizophrenia, bipolar disorder, and major depressive disorder. The Stanley Neuropathology Consortium. *Mol Psychiatry* 2000; **5**: 142–149.

22 Beasley CL, Pennington K, Behan A, Wait R, Dunn MJ, Cotter D. Proteomic analysis of the anterior cingulate cortex in the major psychiatric disorders: Evidence for disease-associated changes. *Proteomics* 2006; **6**: 3414–3425.

23 Martins-de-Souza D, Gattaz WF, Schmitt A, Rewerts C, Maccarrone G, Dias-Neto E *et al.* Prefrontal cortex shotgun proteome analysis reveals altered calcium homeostasis and immune system imbalance in schizophrenia. *Eur Arch Psychiatry Clin Neurosci* 2009; **259**: 151–163.

24 Hensley K, Venkova K, Christov A, Gunning W, Park J. Collapsin response mediator protein-2: an emerging pathologic feature and therapeutic target for neurodisease indications. *Mol Neurobiol* 2011; **43**: 180–191.

25 Yoshimura T, Kawano Y, Arimura N, Kawabata S, Kikuchi A, Kaibuchi K. GSK-3beta regulates phosphorylation of CRMP-2 and neuronal polarity. *Cell* 2005; **120**: 137–149.

26 Deo RC, Schmidt EF, Elhabazi A, Togashi H, Burley SK, Strittmatter SM. Structural bases for CRMP function in plexin-dependent semaphorin3A signaling. *EMBO J* 2004; **23**: 9–22.

27 Kawano Y, Yoshimura T, Tsuboi D, Kawabata S, Kaneko-Kawano T, Shirataki H *et al.* CRMP-2 is involved in kinesin-1-dependent transport of the Sra-1/WAVE1 complex and axon formation. *Mol Cell Biol* 2005; **25**: 9920–9935.

28 Inagaki N, Chihara K, Arimura N, Ménager C, Kawano Y, Matsuo N *et al.* CRMP-2 induces axons in cultured hippocampal neurons. *Nat Neurosci* 2001; **4**: 781–782.

29 Lee H, Joo J, Nah S-S, Kim JW, Kim H-K, Kwon J-T *et al.* Changes in Dpysl2 expression are associated with prenatally stressed rat offspring and susceptibility to schizophrenia in humans. *Int J Mol Med* 2015; **35**: 1574–1586.

30 Liu Y, Pham X, Zhang L, Chen P-L, Burzynski G, McGaughey DM *et al.* Functional variants in DPYSL2 sequence increase risk of schizophrenia and suggest a link to mTOR signaling. *G3 Bethesda Md* 2014; **5**: 61–72.

31 Yamashita R, Suzuki Y, Takeuchi N, Wakaguri H, Ueda T, Sugano S *et al.* Comprehensive detection of human terminal oligo-pyrimidine (TOP) genes and analysis of their characteristics. *Nucleic Acids Res* 2008; **36**: 3707–3715.

32 Morita T, Sobue K. Specification of neuronal polarity regulated by local translation of CRMP2 and Tau via the mTOR-p70S6K pathway. *J Biol Chem* 2009; **284**: 27734–27745.

33 Miloslavski R, Cohen E, Avraham A, Iluz Y, Hayouka Z, Kasir J *et al.* Oxygen sufficiency controls TOP mRNA translation via the TSC-Rheb-mTOR pathway in a 4E-BP-independent manner. *J Mol Cell Biol* 2014; **6**: 255–266.

34 Patursky-Polischuk I, Stolovich-Rain M, Hausner-Hanochi M, Kasir J, Cybulski N, Avruch J *et al.* The TSC-mTOR pathway mediates translational activation of TOP mRNAs by insulin largely in a raptor- or rictor-independent manner. *Mol Cell Biol* 2009; **29**: 640–649.

35 Ng MYM, Levinson DF, Faraone SV, Suarez BK, DeLisi LE, Arinami T *et al.* Meta-analysis of 32 genome-wide linkage studies of schizophrenia. *Mol Psychiatry* 2009; **14**: 774–785.

36 Fallin MD, Lasseter VK, Avramopoulos D, Nicodemus KK, Wolyniec PS, McGrath JA *et al.* Bipolar I disorder and schizophrenia: a 440-single-nucleotide polymorphism screen of 64 candidate genes among Ashkenazi Jewish case-parent trios. *Am J Hum Genet* 2005; **77**: 918–936.

37 Pulver AE, Bale SJ. Availability of schizophrenic patients and their families for genetic linkage studies: findings from the Maryland epidemiology sample. *Genet Epidemiol* 1989; **6**: 671–680.

38 Niranjan TS, Adamczyk A, Bravo HC, Taub MA, Wheelan SJ, Irizarry R *et al.* Effective detection of rare variants in pooled DNA samples using Cross-pool tailcurve analysis. *Genome Biol* 2011; **12**: R93.

39 Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. ESEfinder: a web resource to identify exonic splicing enhancers. *Nucleic Acids Res* 2003; **31**: 3568–3571.

40 Jefferies HB, Fumagalli S, Dennis PB, Reinhard C, Pearson RB, Thomas G. Rapamycin suppresses 5'TOP mRNA translation through inhibition of p70s6k. *EMBO J* 1997; **16**: 3693–3704.

41 Hamilton TL, Stoneley M, Spriggs KA, Bushell M. TOPs and their regulation. *Biochem Soc Trans* 2006; **34**: 12–16.

42 Liu Y. *DPYSL2 and schizophrenia: Identification of sequence variants in DPYSL2 and analysis of their functional significance*. 2010.http://search.proquest.com/dissertations/docview/880574851/abstract/4FEBFF8D671E4054PQ/6 (accessed 7 Jan2016).

43 Fingar DC, Richardson CJ, Tee AR, Cheatham L, Tsou C, Blenis J. mTOR controls cell cycle progression through its cell growth effectors S6K1 and 4E-BP1/eukaryotic translation initiation factor 4E. *Mol Cell Biol* 2004; **24**: 200–216.

44 Ikenoue T, Hong S, Inoki K. Monitoring mammalian target of rapamycin (mTOR) activity. *Methods Enzymol* 2009; **452**: 165–180.

45 Foster KG, Fingar DC. Mammalian target of rapamycin (mTOR): conducting the cellular signaling symphony. *J Biol Chem* 2010; **285**: 14071–14077.

46 Guertin DA, Sabatini DM. Defining the role of mTOR in cancer. *Cancer Cell* 2007; **12**: 9–22.

47 Laplante M, Sabatini DM. mTOR signaling in growth control and disease. *Cell* 2012; **149**: 274–293.

48 Zoncu R, Efeyan A, Sabatini DM. mTOR: from growth signal integration to cancer, diabetes and ageing. *Nat Rev Mol Cell Biol* 2011; **12**: 21–35.

49 Hsieh AC, Liu Y, Edlind MP, Ingolia NT, Janes MR, Sher A *et al.* The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature* 2012; **485**: 55–61.

50 Jaworski J, Sheng M. The growing role of mTOR in neuronal development and plasticity. *Mol Neurobiol* 2006; **34**: 205–219.

51 Gururajan A, van den Buuse M. Is the mTOR-signalling cascade disrupted in Schizophrenia? *J Neurochem* 2014; **129**: 377–387.

52 Swiech L, Perycz M, Malik A, Jaworski J. Role of mTOR in physiology and pathology of the nervous system. *Biochim Biophys Acta* 2008; **1784**: 116–132.

53 Takei N, Nawa H. mTOR signaling and its roles in normal and abnormal brain development. *Front Mol Neurosci* 2014; **7**: 28.

54 Ishizuka K, Kamiya A, Oh EC, Kanki H, Seshadri S, Robinson JF *et al.* DISC1-dependent switch from progenitor proliferation to migration in the developing cortex. *Nature* 2011; **473**: 92–96.

55 Zhu J, Gopinath K, Murali A, Yi G, Hayward SD, Zhu H *et al.* RNA-binding proteins that inhibit RNA virus infection. *Proc Natl Acad Sci U S A* 2007; **104**: 3129–3134.

56 Hu S, Xie Z, Onishi A, Yu X, Jiang L, Lin J *et al.* Profiling the Human Protein-DNA Interactome Reveals ERK2 as a Transcriptional Repressor of Interferon Signaling. *Cell* 2009; **139**: 610–622.

57 Gagnon KT, Maxwell ES. Electrophoretic mobility shift assay for characterizing RNA-protein interaction. *Methods Mol Biol Clifton NJ* 2011; **703**: 275–291.

58 Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* 2013; **8**: 2281–2308.

59 Yang L, Yang JL, Byrne S, Pan J, Church GM. CRISPR/Cas9-Directed Genome Editing of Cultured Cells. *Curr Protoc Mol Biol Ed Frederick M Ausubel Al* 2014; **107**: 31.1.1–31.1.17.

60 Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 2012; **7**: 562–578.

61 Itoh G, Kanno S, Uchida KSK, Chiba S, Sugino S, Watanabe K *et al.* CAMP (C13orf8, ZNF828) is a novel regulator of kinetochore-microtubule attachment. *EMBO J* 2011; **30**: 130–144.

62 Isidor B, Küry S, Rosenfeld JA, Besnard T, Schmitt S, Joss S *et al.* De Novo Truncating Mutations in the kinetochore-microtubules attachment gene CHAMP1 Cause Syndromic Intellectual Disability. *Hum Mutat* 2016. doi:10.1002/humu.22952.

63 Soldati L, Alberoni M, Bianchi G, Canal N, Franceschi M. Erythrocytic calpain-calpastatin system in Alzheimer's disease. *Aging Milan Italy* 1996; **8**: 106–108.

64 Nakayama J, Yoshizawa T, Yamamoto N, Arinami T. Mutation analysis of the calpastatin gene (CAST) in patients with Alzheimer's disease. *Neurosci Lett* 2002; **320**: 77–80.

65 Yuen M, Sandaradura SA, Dowling JJ, Kostyukova AS, Moroz N, Quinlan KG *et al.* Leiomodin-3 dysfunction results in thin filament disorganization and nemaline myopathy. *J Clin Invest* 2014; **124**: 4693–4708.

66 Miyamoto-Sato E, Fujimori S, Ishizaka M, Hirai N, Masuoka K, Saito R *et al.* A comprehensive resource of interacting protein regions for refining human transcription factor networks. *PloS One* 2010; **5**: e9289.

67 Okano HJ, Darnell RB. A hierarchy of Hu RNA binding proteins in developing and adult neurons. *J Neurosci Off J Soc Neurosci* 1997; **17**: 3024–3037.

68 Fukao A, Sasano Y, Imataka H, Inoue K, Sakamoto H, Sonenberg N *et al.* The ELAV protein HuD stimulates cap-dependent translation in a Poly(A)- and eIF4A-dependent manner. *Mol Cell* 2009; **36**: 1007–1017.

69 Deschênes-Furry J, Perrone-Bizzozero N, Jasmin BJ. The RNA-binding protein HuD: a regulator of neuronal differentiation, maintenance and plasticity. *BioEssays News Rev Mol Cell Dev Biol* 2006; **28**: 822–833.

70 Sosanya NM, Cacheaux LP, Workman ER, Niere F, Perrone-Bizzozero NI, Raab-Graham KF. Mammalian Target of Rapamycin (mTOR) Tagging Promotes Dendritic Branch Variability through the Capture of Ca2+/Calmodulin-dependent Protein Kinase II α (CaMKIIα) mRNAs by the RNA-binding Protein HuD. *J Biol Chem* 2015; **290**: 16357–16371.

71 Shaw G, Morse S, Ararat M, Graham FL. Preferential transformation of human neuronal cells by human adenoviruses and the origin of HEK 293 cells. *FASEB J Off Publ Fed Am Soc Exp Biol* 2002; **16**: 869–871.

72 Ho S-Y, Chao C-Y, Huang H-L, Chiu T-W, Charoenkwan P, Hwang E. NeurphologyJ: an automatic neuronal morphology quantification method and its application in pharmacological discovery. *BMC Bioinformatics* 2011; **12**: 230.

73 Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014; **511**: 421–427.

74 Pennington K, Dicker P, Dunn MJ, Cotter DR. Proteomic analysis reveals protein changes within layer 2 of the insular cortex in schizophrenia. *PROTEOMICS* 2008; **8**: 5097–5107.

75 Martins-de-Souza D, Schmitt A, Röder R, Lebar M, Schneider-Axmann T, Falkai P *et al.* Sex-specific proteome differences in the anterior cingulate cortex of schizophrenia. *J Psychiatr Res* 2010; **44**: 989–991.

76 English JA, Pennington K, Dunn MJ, Cotter DR. The Neuroproteomics of Schizophrenia. *Biol Psychiatry* 2011; **69**: 163–172.

77 Chen J, Lin M, Hrabovsky A, Pedrosa E, Dean J, Jain S *et al.* ZNF804A Transcriptional Networks in Differentiating Neurons Derived

from Induced Pluripotent Stem Cells of Human Origin. *PloS One* 2015; **10**: e0124597.

78 Tao R, Cousijn H, Jaffe AE, Burnet PWJ, Edwards F, Eastwood SL *et al.* Expression of ZNF804A in human brain and alterations in schizophrenia, bipolar disorder, and major depressive disorder: a novel transcript fetally regulated by the psychosis risk variant rs1344706. *JAMA Psychiatry* 2014; **71**: 1112–1120.

79 Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ *et al.* The Connectivity Map: Using Gene-Expression Signatures to Connect Small Molecules, Genes, and Disease. *Science* 2006; **313**: 1929–1935.

80 Lamb J. The Connectivity Map: a new tool for biomedical research. *Nat Rev Cancer* 2007; **7**: 54–60.

81 Hay N, Sonenberg N. Upstream and downstream of mTOR. *Genes Dev* 2004; **18**: 1926–1945.

82 Efeyan A, Sabatini DM. mTOR and cancer: many loops in one pathway. *Curr Opin Cell Biol* 2010; **22**: 169–176.

83 St Clair D, Xu M, Wang P, Yu Y, Fang Y, Zhang F *et al.* Rates of adult schizophrenia following prenatal exposure to the Chinese famine of 1959-1961. *JAMA* 2005; **294**: 557–562.

84 Susser ES, Lin SP. Schizophrenia after prenatal exposure to the Dutch Hunger Winter of 1944-1945. *Arch Gen Psychiatry* 1992; **49**: 983–988.

85 Land SC, Tee AR. Hypoxia-inducible factor 1alpha is regulated by the mammalian target of rapamycin (mTOR) via an mTOR signaling motif. *J Biol Chem* 2007; **282**: 20534–20543.

86 Howell JJ, Manning BD. mTOR couples cellular nutrient sensing to organismal metabolic homeostasis. *Trends Endocrinol Metab TEM* 2011; **22**: 94–102.

87 Wesseling H, Rahmoune H, Tricklebank M, Guest PC, Bahn S. A targeted multiplexed proteomic investigation identifies ketamine-induced changes in immune markers in rat serum and expression changes in protein kinases/phosphatases in rat brain. *J Proteome Res* 2015; **14**: 411–421.

88 Zhao X, Tang R, Xiao Z, Shi Y, Feng G, Gu N *et al.* An investigation of the dihydropyrimidinase-like 2 (DPYSL2) gene in schizophrenia: genetic association study and expression analysis. *Int J*

Neuropsychopharmacol Off Sci J Coll Int Neuropsychopharmacol CINP 2006; **9**: 705–712.

89 Khandaker GM, Cousins L, Deakin J, Lennox BR, Yolken R, Jones PB. Inflammation and immunity in schizophrenia: implications for pathophysiology and treatment. *Lancet Psychiatry* 2015; **2**: 258–270.

90 Avramopoulos D, Pearce BD, McGrath J, Wolyniec P, Wang R, Eckart N *et al.* Infection and inflammation in schizophrenia and bipolar disorder: a genome wide study for interactions with genetic variation. *PloS One* 2015; **10**: e0116696.

91 Cash-Padgett T, Jaaro-Peled H. DISC1 mouse models as a tool to decipher gene-environment interactions in psychiatric disorders. *Front Behav Neurosci* 2013; **7**: 113.

92 Deussing JM. Targeted mutagenesis tools for modelling psychiatric disorders. *Cell Tissue Res* 2013; **354**: 9–25.

93 O'Tuathaigh CMP, Desbonnet L, Waddington JL. Genetically modified mice related to schizophrenia and other psychoses: seeking phenotypic insights into the pathobiology and treatment of negative symptoms. *Eur Neuropsychopharmacol J Eur Coll Neuropsychopharmacol* 2014; **24**: 800–821.

94 Sarnyai Z, Jashar C, Olivier B. Modeling combined schizophrenia-related behavioral and metabolic phenotypes in rodents. *Behav Brain Res* 2015; **276**: 130–142.

95 Seong E, Seasholtz AF, Burmeister M. Mouse models for psychiatric disorders. *Trends Genet TIG* 2002; **18**: 643–650.

96 Simpson EH, Waltz JA, Kellendonk C, Balsam PD. Schizophrenia in translation: dissecting motivation in schizophrenia and rodents. *Schizophr Bull* 2012; **38**: 1111–1117.

97 Papaleo F, Lipska BK, Weinberger DR. Mouse models of genetic effects on cognition: relevance to schizophrenia. *Neuropharmacology* 2012; **62**: 1204–1220.

98 Bork K, Karbe Y, Pollscheit J, Glaubitz N, Nöhring S, Horstkorte R. Role of collapsin response mediator protein-2 in neurite outgrowth of PC12 cells. *Neuroreport* 2010; **21**: 641–645.

99 Zevnik B, Uyttersprot NC, Perez AV, Bothe GWM, Kern H, Kauselmann G. C57BL/6N albino/agouti mutant mice as embryo

donors for efficient germline transmission of C57BL/6 ES cells. *PloS One* 2014; **9**: e90570.

100 Ryder E, Wong K, Gleeson D, Keane TM, Sethi D, Vyas S *et al.* Genomic analysis of a novel spontaneous albino C57BL/6N mouse strain. *Genes N Y N 2000* 2013; **51**: 523–528.

101 Yokoyama T, Silversides DW, Waymire KG, Kwon BS, Takeuchi T, Overbeek PA. Conserved cysteine to serine mutation in tyrosinase is responsible for the classical albino mutation in laboratory mice. *Nucleic Acids Res* 1990; **18**: 7293–7298.

102 Crawley JN. Exploratory behavior models of anxiety in mice. *Neurosci Biobehav Rev* 1985; **9**: 37–44.

103 Ayhan Y, Abazyan B, Nomura J, Kim R, Ladenheim B, Krasnova IN *et al.* Differential effects of prenatal and postnatal expressions of mutant human DISC1 on neurobehavioral phenotypes in transgenic mice: evidence for neurodevelopmental origin of major psychiatric disorders. *Mol Psychiatry* 2011; **16**: 293–306.

104 Lalonde R. The neurobiological basis of spontaneous alternation. *Neurosci Biobehav Rev* 2002; **26**: 91–104.

105 Adamczyk A, Mejias R, Takamiya K, Yocum J, Krasnova IN, Calderon J *et al.* GluA3-deficiency in mice is associated with increased social and aggressive behavior and elevated dopamine in striatum. *Behav Brain Res* 2012; **229**: 265–272.

106 Conrad CD, Galea LA, Kuroda Y, McEwen BS. Chronic stress impairs rat spatial memory on the Y maze, and this effect is blocked by tianeptine pretreatment. *Behav Neurosci* 1996; **110**: 1321–1334.

107 Conrad CD, Lupien SJ, Thanasoulis LC, McEwen BS. The effects of type I and type II corticosteroid receptor agonists on exploratory behavior and spatial memory in the Y-maze. *Brain Res* 1997; **759**: 76–83.

108 Belzung C, Griebel G. Measuring normal and pathological anxiety-like behaviour in mice: a review. *Behav Brain Res* 2001; **125**: 141–149.

109 Pinheiro SH, Zangrossi H, Del-Ben CM, Graeff FG. Elevated mazes as animal models of anxiety: effects of serotonergic agents. *An Acad Bras Ciênc* 2007; **79**: 71–85.

110 Pletnikov MV, Ayhan Y, Nikolskaia O, Xu Y, Ovanesov MV, Huang H *et al.* Inducible expression of mutant human DISC1 in mice is associated with brain and behavioral abnormalities reminiscent of schizophrenia. *Mol Psychiatry* 2008; **13**: 173–186, 115.

111 Schmidt EF, Strittmatter SM. The CRMP family of proteins and their role in Sema3A signaling. *Adv Exp Med Biol* 2007; **600**: 1–11.

Curriculum Vitae

Xuan T. Pham
xpham2@jhmi.edu

EDUCATION

2011 – 2016 **Johns Hopkins University,** Baltimore, MD
Ph.D., Human Genetics and Molecular Biology

Advisor: Dimitrios Avramopoulos

2005 – 2009 **Virginia Commonwealth University,** Richmond, VA
B.S. Biology *Magna Cum Laude*

RESEARCH EXPERIENCE

2011 – 2016 **Doctoral thesis**, Johns Hopkins University, Baltimore, MD
Characterized a functional variant of *DPYSL2* in schizophrenia risk using molecular, *in vitro* and *in vivo* model systems

Principle Investigator: Dr. Dimitrios Avramopoulos

2008 – 2009 **Undergraduate Research Assistant**, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA

Principle Investigator: Dr. Jack Hettema
Elucidated the association study between GABA receptor genes and anxiety spectrum disorders

2007 – 2008 **Undergraduate Research Assistant**, Department of Infectious Disease, Virginia Commonwealth University, Richmond, VA

Principle Investigator: Dr. Adrianna Rosato
Investigated the risk factors and prevalence of Methicillin-Resistant *S. Aureus* in the community

2006 – 2007 **Undergraduate Research Assistant**, Department of Anatomy and Neurobiology, Virginia Commonwealth University, Richmond, VA

Principle Investigator: Dr. Kimberle Jacobs
Studied neuronal hyperexcitability in developmental epilepsy through the application of a gap-junction blocker in a rat model of intractable epilepsy.

INDUSTRY EXPERIENCE

2010 – 2011 **Lead Lab Technician,** American International Biotechnology, Clinical Genetics Department, Richmond, VA

Principle Investigator: Dr. Sarah Helber
Performed research, developed, and optimized new clinical assays for the detection and diagnosis of sexually transmitted diseases, cystic fibrosis, and prostate cancer

PUBLICATIONS

Liu, Y., **Pham, X.**, Zhang, L., Chen, P., Burzynski, G., McGaughey, D.M., He, S., McGrath, J.A., Wolyniec, P., Fallin, M.D., Pierce, M.S., McCallion, A.M., Pulver, A.E., Avramopoulos, D., and Valle, D. "Functional Variants in *DPYSL2* Sequence Increase Schizophrenia Risk and Suggest a Link to mTOR Signaling." *Genes Genomes Genetics.* 2015; 5(1): 61-72.

Pham, X., C. Sun, van den Oord, E., Chen, X., Neale, M.C., Kendler, K.S., and Hettema, J.M. "Association study between GABA receptor genes and anxiety spectrum disorders." *Depression and Anxiety.* 2009; 26: 998-1003.

ADDITIONAL WRITING EXPERIENCE

2015 – Present Science Writer & Webinar Moderator**, LabRoots.com**

2015 – Present Science Writer, **OMIM.org**

2014 – Present Founder & Writer**, TheGeneTwist.com**

2015 – 2016 Science Writer
biomedicalodyssey.blogs.hopkinsmedicine.org

PRESENTATIONS

October 2015 **65th Meeting of the American Society of Human Genetics**, Baltimore, MD
"Schizophrenia associated variation in DPYSL2 perturbs mTOR signaling and produces cellular phenotypes."
Platform talk, **Epstein Trainee Award for Excellence in Human Genetics Research – Semifinalist**

October 2014 **64th Meeting of the American Society of Human Genetics,** San Diego, CA
"A polymorphic di-nucleotide repeat (DNR) variant in the 5'UTR of DPYSL2 gene affects its regulation by mTOR," poster

April 2009	**Honors College Undergraduate Research Symposium,** Richmond, VA "Association study between GABA receptor genes and anxiety spectrum disorders," poster
August 2008	**Leadership Alliance National Symposium,** Hartford, CT "Sequence analysis on a region of chromosome 10q22-23 associated with the delusional schizophrenia phenotype," poster
April 2007	**National Conference on Undergraduate Research,** San Francisco, CA "Hypersynchrony and effects of Mefloquine in maldeveloped epileptogenic neocortex," poster

HONORS & DISTINCTIONS

2015	Semifinalist: Charles J. Epstein Trainee Award for Excellence in Human Genetics Research
2009	Outstanding Biology Senior Award, VCU
2008	Research Stipend, Leadership Alliance
2008	Honors College Study Abroad Grant, VCU
2007	Honors College Travel Grant, VCU
2005-2009	Dean's Award - Excellence in Academics, VCU
2005-2009	Virginia Urology Center Merit Scholarship
2005-2009	Presidential Scholar, VCU

LEADERSHIP ROLES

2014 – 2015	**Visiting Scientist,** "Genome Geeks Are In!" Program, Smithsonian Museum of Natural History, Washington, DC
2013 – 2015	**Mentor to undergraduate students,** Psychiatry Summer Training and Research, Johns Hopkins University, Baltimore, MD
2005 – 2008	**Emergency Medical Technician-Basic Provider**, Manchester Volunteer Rescue Squad, Richmond, VA